Detecting the necessity of thinnings with deep learning

Philipp Satlawa

Master of Science

School of Informatics
University of Edinburgh

2021

Abstract

Timely information about the necessity of thinning in the forest is vital for forest management to maintain a healthy forest while maximising income. Currently, very high spatial resolution remote sensing data can provide crucial assistance for the experts to evaluate the maturity of thinnings. Yet, this task is still predominantly determined in the field and demands extensive resources, thus causing high temporal resolution in the procurement. In this study, we propose to employ a deep convolutional neural network (DCNN) to detect the necessity and urgency of thinnings by using only remote sensing data. Notably, we merge very high spatial resolution RGB and near-infrared orthophotos, canopy height model (CHM), digital terrain model (DTM), slope and the reference data into one data set to train the DCNN. Experts acquired the reference data in spruce dominated forests in the Austrian Alps. After tuning the model's hyper-parameters on the data set, the model achieves a test set F1 score of 82.23%. Consequently, we conclude that DCNNs are indeed capable of detecting the need for thinning in forests. In contrast, all attempts of assessing the urgency of thinnings with DCNNs proved to be unsuccessful. However, additional data such as age or yield class has the potential of improving the results. We further investigate the influence of the individual input features on the model performance. For example, orthophotos appear to contain the most relevant information for detecting the demand for thinning. Moreover, we observe a gain in performance by adding CHM and slope, whereas adding the DTM harms the model's performance.

Acknowledgements

First and foremost, I want to thank my supervisor, Bob Fisher, for guiding me through the entire project as well as stimulating new inspiring ideas and research directions. I am also very grateful to the Austrian Federal Forests (ÖBF AG) for supporting the study and providing the necessary data. Finally, I would like to thank my family for the unconditional support they gave me during my academic journey, especially Paula Ruiz Rodrigo and Sylwia Whitman, for providing valuable feedback on this dissertation.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Philipp Satlawa*)

Table of Contents

Chapter 1

Introduction

Maintaining a healthy, stable forest to produce valuable wood requires fostering the forest. An essential technique to accomplish that is a silvicultural operation called thinning (Daume and Robertson (2000)). The main objective of thinning is to regulate the vertical space of trees, thus steering the allocation of the available resources (e.g. sunlight, water, nutrients) into the stems of remaining higher quality trees (Mitchell (2000)). Although the primary objective of thinning is to prepare the forest stand for the final harvest, it also provides forest owners with the opportunity of obtaining additional income by selling the removed trees.

Determining if a forest stand needs thinning is a complex task since it is dependent on many factors such as soil quality, age and tree species composition (Juodvalkis et al. (2005)). This assessment makes assessing a forest stand for the necessity of thinning a non-trivial task and is why the job is still mainly conducted by specialised forest personnel. Sending personnel into the field is often very expensive. Hence, thinning assessment is either not done at all like in the case of small forest owners or done at long time intervals, which might be sufficient for slowly growing sites, however vital areas with high mean annual increment are often overlooked, which results in a sub-optimal timing of thinning in these stands.

In recent times, the proliferation in the amount and enhancement in the quality of remote sensing data has raised the possibility of providing detailed information over large areas from above (Ghamisi et al. (2017)). Furthermore, this data is currently utilised as a base for everyday tasks such as planning felling operations and conducting nature protection projects. Therefore, remote sensing imagery is now an essential tool for operational forest management.

Although assessment of the necessity of thinning is not done exclusively using

remote sensing data, it provides essential information for the evaluation. In particular, colour-infrared (CIR) orthophotos provide insight into tree species composition, crown width, crown density, and all-important parameters for determining the necessity of thinning. In addition, another crucial parameter for assessing thinnings is the top height of the forest stand that can be estimated through a remote sensing product named the canopy height model (CHM).

Furthermore, we can employ the data to develop models capable of automatic extraction of valuable information. This processing is possible due to recent increases in computing power that allows more capable machine learning (ML) algorithms to be deployed. This advancement resulted in the development of ML models, such as classifying tree species (Immitzer et al. (2019)) or estimating the standing volume (Halme et al. (2019)).

Nevertheless, little research has been done to detect the necessity of thinning solely with remote sensing data. That is why our main target for this study is to develop a model capable of achieving this task. An accurate prediction of thinnings from remote sensing data has the potential of delivering information for vast forest areas promptly and thus improving the stability and wood quality of forests. The main objectives of this study are:

- To evaluate the possibility of detecting the need of thinning with deep convolutional neural networks (DCNNs) trained on very high spatial resolution imagery.
- To assess the possibility of additional differentiation between thinnings with different urgency.
- To identify the main sources of errors
- To investigate the importance of the individual data inputs.

Based on a study of the literature, we concluded that no researchers have previously explored using DCNNs to estimate the need for thinning (see Chapter 2). We developed a machine learning based approach (see Chapter 4) that used data from the Austrian Federal Forest system (see Chapter 3) which was then preprocessed (see Section 4.1) to make it suitable for a deep-learning based classifier (see Section 4.2). The trained classifier was capable of recognising when thinning was needed with an accuracy of 82.5% (see Chapter 6). An overview of the data prepocessing is displayed in Figure 4.1, the training approach can be seen in Figure 4.3, and an example of aerial forest data and its classifications is seen here in Figure 5.2).

Chapter 2

Related work

This chapter provides background information related to detecting the necessity of thinnings with the help of remote sensing. First, we introduce thinning, its effect on the forest stand, and its impact on the timber quality. After that, we show the current state of remote sensing in forestry on its most important research questions. Finally, we overview the current state of the art on semantic segmentation techniques and their application in remote sensing.

2.1 Thinning

Thinning can be seen as the primary steering technique in preparation for the final harvest. Although many types of thinning have emerged through the history of forestry, the main objective of all of them is to remove some trees and accelerate growth in the remaining future crop trees. Notably, this enables the remaining trees to allocate the newly available resources mainly into their basal stem (Mitchell (2000), Holgén et al. (2003)), which is economically the most valuable part of the tree.

Furthermore, by applying selective thinning, hence selecting the most promising future crop trees, the log quality can be significantly enhanced (Stirling et al. (2000), Macdonald et al. (2010)). Thus, thinning provides an opportunity to increase the amount of good quality timber in the final harvest while producing additional income from the sale of the removed trees. Since the main objective of many forest owners is to maximise the net present value of their forest area, thinned stands outperform unmanaged stands in that matter (Spellmann and Schmidt (2003), Hynynen et al. (2005), Hein et al. (2008)).

One of the main concerns of performing thinning is the enhanced risk for damages

to the forest due to high wind or snow. Various studies have shown that right after thinning, the forest stand's stability is lower due to the higher roughness of the tree crowns (Persson (1975)). Although this is true for the time right after thinning, the risk of injuries decreases with time until no additional risk is present anymore (Persson (1975), MacKenzie (1976)). Nonetheless, thinning in Norway spruce stands has been recorded to reduce the h/d ratio of trees (height to diameter ratio) (Slodicak et al. (2005)). The h/d ratio is an indicator for the individual stability of a tree, where low h/d values are equivalent to high tree stability (Pollanschutz (1980), Valinger and Fridman (1999), Valinger et al. (2006)). Thus, newer research suggests that early heavy thinning might even increase the stability of a stand (Schütz et al. (2006)). Thereby, timeliness is crucial, as delaying thinnings results in a risk increase of damage in the stands (Pollanschutz (1980), Cameron (2002)).

2.2 Remote sensing in forestry

Utilising remote sensing imagery for extracting information about the forest structure and its employment in forest planning has been practised since the 1960s (Avery (1966)). Remote sensing provides the opportunity to gather uniform information about considerable areas otherwise impossible to collect from field measurements. Automated derivation of crucial information from the forest, such as tree species classification and wood volume estimation, has long been a goal of researchers in forestry (Kangas et al. (2018)). Traditionally, collecting information about the forest structure was based on sending experts into the field as in conventional forest inventories. This process is costly and often not affordable for small forest owners. With the advances in sensor technology as well as computing power, the interest in applying algorithms for automated extraction of forest parameters from remotely sensed spectral information increased (Ma et al. (2019)).

At present, the leading sensing instrument technologies applied for retrieving forestry relevant metrics are multi-spectral or hyperspectral cameras, light detection and ranging (lidar), and to a certain extent, Synthetic-aperture radar (SAR) (Holopainen et al. (2014)). These instruments can be mounted on three different platforms, satellite, aircraft and unmanned aerial vehicle (UAV), each giving the sensing instrument different ranges of spatial resolution. Moreover, sensors mounted on a plane can support conventional forest inventories and provide helpful information for operational forest management (Magnussen et al. (2018), Kangas et al. (2018)).

Despite the importance of forest operations such as thinning, few studies have tried to create models which can predict the necessity of thinning directly from remote sensing data (Vastaranta et al. (2011)). Most research has been focused on change detection, tree species classification as well as wood volume estimation. Although substantial progress has been made in all the introduced research questions, they are far from being solved and still active research subjects.

2.2.1 Change detection

The forest changes continuously, whether through carried out forest operations such as thinning or clear-cutting or by forest damages caused by heavy winds, fire and other natural disasters. The man-made changes can be reported and easily updated. However, the changes induced by nature must be spotted differently. Automatic detection of changes in the forest using aerial imagery shows promising results in areas where severe changes happened, such as clear-cutting and intense storm damage. However, moderate changes such as thinnings were much harder to discover Hyvönen et al. (2010)). When using airborne laser scanning (ALS) as the data source to detect changes, Yu et al. (2004) were able to identify the removal of individual trees. Nonetheless, acquiring aerial images or even more expensive ALS data is often done at considerable intervals. Hence, a more interesting remote sensing platform for this problem are satellites. Due to their higher temporal resolution, satellites can provide much more timely data as it is often required in disaster response. In particular, SAR data can provide data even on overcast days. Although weather conditions influence the signal, this noise can be filtered, as Olesk et al. (2015) has demonstrated. Thereby they were able to detect changes in forest areas greater than 1 ha with Sentinel-1 data.

2.2.2 Tree species

Accurate assessment of tree species composition over large areas is easily possible by utilising remote sensing data. However, automatic classification of tree species with remote sensing data has proven to be a challenging task (Fassnacht et al. (2016)). However, recent studies have shown great potential either by using hyperspectral imagery (Ballanti et al. (2016), Fricker et al. (2019)) from one acquisition or by utilising multi-temporal and multi-spectral data (Immitzer et al. (2019), Axelsson et al. (2021)). The latter is especially impressive since Immitzer et al. (2019) employed just freely available Sentinel-2 imagery, resulting in very high classification accuracy. Yet, the spatial

resolution of the Sentinel-2 optical sensor is limited to 10 m (Bands 2,3,4,8). Using the WorldView-3 satellite with a spatial resolution of 1.6 m of the Bands RGB and near-infrared (NIR) as well as a deep convolutional neural networks (DCNN), Yan et al. (2021) has been able to achieve even more refined individual tree classification.

2.2.3 Tree height and wood volume

Besides tree species composition, the other two most essential forest attributes for forest managers are tree heights and stocking wood volume. Stocking wood volume is highly correlated with tree heights since tree heights are, besides the basal area, the most critical parameter for estimating wood volume. Therefore many studies that derive the tree heights from remote sensing data provide results on both forest parameters.

Up until now, research has deployed multi-spectral and hyperspectral cameras, lidar as well as SAR to determine wood volume with varying success. Deriving tree heights from airborne hyper-spectral data with over a hundred bands ranging from the visible to the near-infrared is feasible (RMSE 6.39 m). Nonetheless, a similar accuracy can be achieved by employing just multi-spectral satellite data (RMSE 6.14 m) (Halme et al. (2019), Cooper et al. (2021)). Additional input data can further improve the performance of models determining tree heights from remote sensing data as in the research of Liu et al. (2019), who fused SAR (Sentinel-1), multi-spectral (Sentinel-2) and digital elevation model (DEM) data successfully to estimate the mean height to obtain an RMSE of 2.9 m.

It has been shown that on comparable spatial resolutions, lidar, which captures directly three-dimensional information, is more precise in estimating tree heights and thus wood volume than any other commonly used sensor technology (Bohlin et al. (2017), Ganz et al. (2019)). When lidar is mounted on a UVA, the captured data even surpasses manual field measurements of tree heights in precision (Ganz et al. (2019)) (RMSE 0.43 m). However, the drawback is the higher operational cost of the sensor and hence its lower temporal resolution. Although spaceborne lidar devices exist that can resolve the temporal resolution problem, the laser beam is sparse and needs to be fused with additional data to provide canopy heights covering the entire area (RMSE 3.4 m) (Boudreau et al. (2008), Simard et al. (2011)).

A suitable alternative to lidar is the estimation of tree heights by applying stereophotogrammetry on aerial imagery. Contrary to the more expensive ALS acquisition, many countries have annual schedules for aerial image acquisitions. For instance, Austria

and Sweden renew one-third of their area every year. Moreover, Bohlin et al. (2017) demonstrated that the canopy heights derived from aerial imagery by applying photogrammetry deliver comparable accuracies (RMSE 1.6 m) similar to those obtained by lidar.

2.2.4 Forest operations

Despite the high importance of planning forest operations timely, few studies have been conducted to predict thinnings from remote sensing data. Just two known studies addressed the challenge. Hyvönen (2002) tried to predict forest operations at the stand-level by using Landsat-TM satellite imagery with moderate success. A second study was carried out by Vastaranta et al. (2011) to predict the thinning maturity at the stand-level from ALS data derived features. This study showed much better results with classification accuracy ranging from 79% to 83% for predicting the timing of the subsequent thinning. Nevertheless, ALS data is still expensive to obtain and, in many countries, not systematically acquired. Another approach to predict the necessity of thinning is to utilise key forest parameters acquired through remote sensing together with additional inventory data to create statistical models (Haara and Korhonen (2004)).

2.3 Deep learning architectures for semantic segmentation

Semantic segmentation is a computer vision task where the algorithm labels each pixel or patch of an image to a predefined range of classes. Accelerated by the first end-to-end fully convolutional network (FCN) (Long et al. (2015)), the utilisation of end-to-end DCNNs for the task of semantic segmentation increased strongly. Since then, many different architectures were proposed and adapted to remote sensing applications (Volpi and Tuia (2018), Yue et al. (2019), Diakogiannis et al. (2020)). The main advantage of using DCNNs for semantic segmentation is their effectiveness in extracting complex features from wide receptive fields. Nonetheless, this capability comes with the price of not being able to maintain high spatial resolution and results in inaccurate and blurred boundaries between the classes. To counteract this, newer DCNNs use more pronounced/distinct encoder-decoder architectures with skip connections. For example, UNet applies such skip connections in its symmetrical encoder-decoder ar-

chitecture (Ronneberger et al. (2015)), where the features extracted in the encoder are directly coupled to the corresponding decoder layers. The outstanding performance results, together with the simplicity of the architecture, ensured the wide adoption in the remote sensing research community for a variety of applications such as ship detection (Hordiiuk et al. (2019)), road extraction (Chen et al. (2021)) and land cover classification (Stoian et al. (2019)).

To obtain models that can learn more complex input representations, deeper DCNNs with more stacked layers were created. However, these deeper DCNNs often resulted in worse-performing models than the more shallow predecessors due to the degradation problem (He and Sun (2015)). This issue was resolved by Deep residual networks (ResNet) that employed residual blocks, which added an identity shortcut connection and overcame the degradation problem (He et al. (2016)). Further development of the DCNNs resulted in the creation of a network architecture called DenseNet, which utilises dense blocks that introduce direct connections from any layer to all subsequent layers to improve further the information flow between layers (Huang et al. (2017)). Jegou et al. (2017) adopted the DenseNet connection structure for semantic segmentation by applying dense blocks to the UNet like symmetric encoder-decoder structure. This merge resulted in a DCNN called FC-DenseNet that needs fewer parameters while performing better on various semantic segmentation challenges.

Another compelling approach to resolve the trade-off between high context extraction with heavy downsampling and accurate boundary prediction is the use of dilated or atrous convolutions. Atrous convolutions contribute with a convolution filter that is spaced apart, thus attaining a wider field of view while retaining its spatial dimension. Chen et al. (2014) proposed a DCNN architecture called DeepLabv1 that incorporates the atrous convolutions in a VGG-16 architecture to address the trade-off between high context extraction with heavy downsampling. After several revisions of this architecture that included, among other things, a Spatial Pyramidal Pooling that was introduced in SPPNet (He et al. (2015b)) and adopted by Chen et al. (2018a) to create the Atrous Spatial Pyramid Pooling (ASPP) in DeepLabv2. The current network architecture DeepLabv3+ (Chen et al. (2018b)), provides some of the best results in semantic segmentation challenges. Furthermore, its use in remote sensing applications is auspicious. For example, Liu et al. (2021) showed the network's excellent performance in classifying marsh vegetation in China. Accordingly, we will use DeepLabv3+ as the network used in Chapter 4 to classify tree thinning regions.

Chapter 3

Materials

For training the deep net, we used data collected for the Lungau region through aerial imagery, airborne laser scanning (ALS) as well as data from forest management plans. This data was first cleaned and preprocessed before being used as input data for the machine learning algorithms (ingested into the ml algorithms to train the models), resulting in 5 distinct input types as illustrated in Table 3.1. This chapter defines what kind of thinning is used in this study, describes the data and its acquisition, and the data preprocessing.

3.1 Study area

Located in the Lungau region (Tamsweg district), Austria (47°00' - 47°13'N, 13°23' - 14°0'E, UTM/WGS84 projection), the study area is managed by the Austrian Federal Forests (ÖBF AG). As part of the Central Eastern Alps, the area is mountainous with an altitude range 993 – 1906 m and has a distinct alpine climate with an average annual temperature of 5.2 °C, mean annual precipitation ranging between 770 – 840 mm and a snow cover minimum of 1 cm on 105 days per year. The study area illustrated in Figure 3.1 has an area of 21826.55 ha and is predominantly covered with forest (63.9% of the area). The forest area is further divided into commercial forest and protective forest. The commercial forest is managed to maximise the income from timber production while minimising the risk of forest damage. In contrast, a protective forest objective is to protect against avalanches, rockfall, erosion and floods. Although thinnings are planned in both forest types, thinnings in the protective forest aim to ensure its protective function, whereas thinnings, as specified in section 3.2, are just planned in a commercial forest. We are merely interested in the part where commercial thinnings

are feasible. Hence we restrict the study area to the commercial forest. The commercial forest has an extent of 9353.54 ha and is stocked with mainly coniferous forest. The most frequent tree species are 81.0% Norway spruce (*Picea abies*, 81.0%) and European larch (*Larix decidua*, 17.6%). The remaining 1.4% of the area are stocked with deciduous forest.

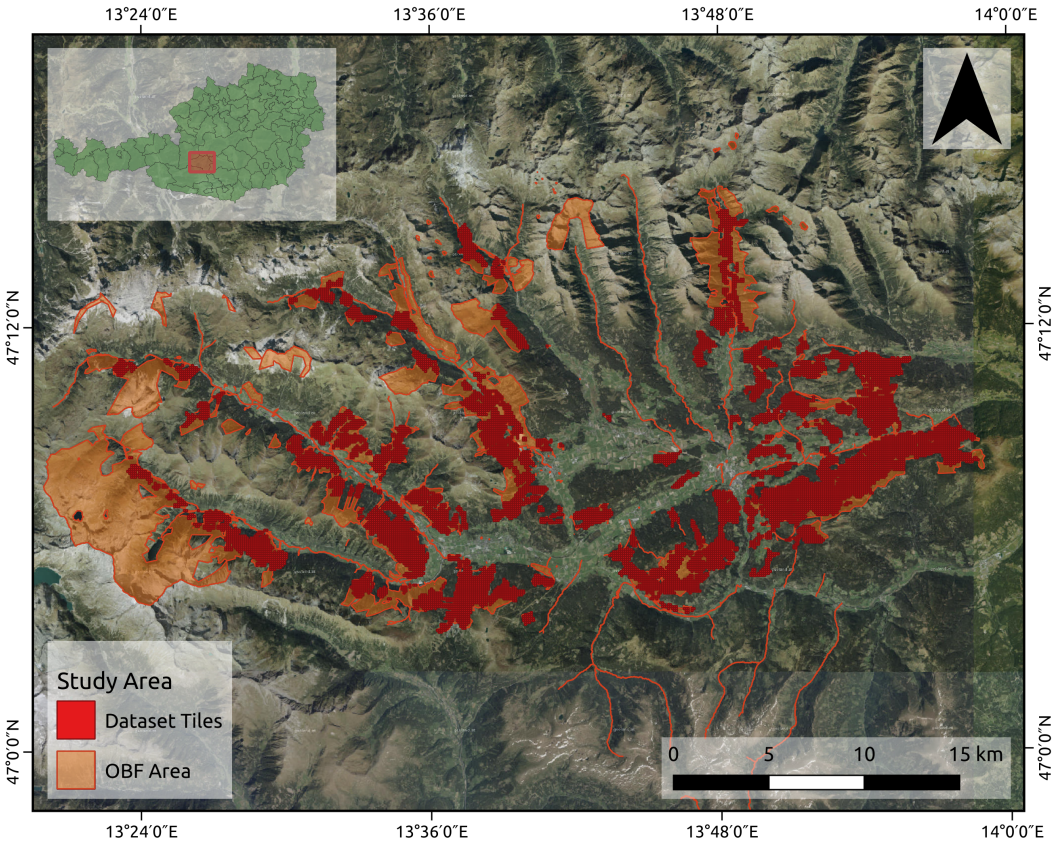


Figure 3.1: Study area of Lungau, Austria. Austrian Federal Forest area is coloured in orange, includes commercial and protective forest as well as non forest areas. Shown in red are the tiles that are used to create the data set. The background image is a true colour orthophoto from airborne photography (basemap (2021)).

3.2 Thinnings

As discussed in section 2.1, thinnings are a crucial technique to foster forest stands. However, optimal thinning schemes differ by the tree species present, the previously executed measures, as well as the density of the standing trees. Due to this variation, it is essential to define the type of thinning that is being used throughout this study.

The study area is predominantly stocked with Norway spruce dominated coniferous forest. Abetz (1970) and later Hein et al. (2008) found that the highest economic return on investment is produced by selective crown thinning with a selection of target trees for this typical forest type. Furthermore, the Austrian Federal Forest adopted this thinning type as the standard thinning scheme and thus, nearly all thinnings are planned as such. Therefore, when we refer to thinning in this study, it is equivalent to this specific thinning type. In practical terms, thinnings in Norway Spruce stands of the Austrian Federal Forest are executed 2-3 times during a lifetime of a stand, where 1/4 to 1/3 of the actual stocking volume is being harvested each time. The first thinning is carried out typically at around 13-19 m top height, followed by a second thinning at approximately 20-30 m top height and a third thinning on good-growing sites. The exact timing of the procedure depends on the density of the standing trees, which is related to conditions such as the timing of the previous thinning and the growth performance on the site.

3.3 Data acquisition

As shown in Table 3.1, all data came from three sources and was acquired at three disparate points in time.

Table 3.1: Data sources. NIR: near infrared, CHM: Canopy height model, DTM: digital train model, AI: aerial imagery, ALS: airborne laser scanning, RD: reference data, Res.: spatial resolution, S.E.: standard error and Year: acquisition year.

Name	Source	Format	Bands	Res. [m]	S.E. [m]	Year
RGB + NIR	AI	GeoTIFF	4	0.2	-	2018
CHM	AI	GeoTIFF	1	1	1	2018
DTM	ALS	GeoTIFF	1	1	0.15	2013
Slope	ALS	GeoTIFF	1	1	-	2013
Ground Truth	RD	Shape	3	-	-	2017

3.3.1 Aerial imagery

The acquisition of the aerial imagery is performed by the States of Austria, where one-third of Austria is updated every given year, hence every three years, the whole of Aus-

tria is updated. All aerial imagery for the study site was recorded during the two days 11.09.2018, and 12.09.2018 with an UltraCam Eagle Mark 3 431S61680X916102-f100 mounted on a Beechcraft Super King Air B200 D-IWAW. Four bands were acquired with a spatial resolution of 20 cm, the three bands for RGB and one band in the near-infrared (NIR) spectrum (Table 3.1). After the acquisition, the raw data was geometrically and radiometrically corrected with the corresponding calibration data of the camera and the four channels were stitched together using the monolithic stitching method (Gruber et al. (2012)). The second product that was derived from the aerial photos is the canopy height model (CHM). To obtain the CHM, first, the digital surface model (DSM) is calculated using photogrammetry from overlapping air images. Then, the CHM is calculated by subtracting the DSM from the DTM. For this study, the Federal Forest Office (BFW) calculated the CHM.

3.3.2 Airborne laser scanning

Acquisition of the utilised digital terrain model (DTM) was performed by airborne laser scanning (ALS) in 2013 as part of the EU project INTERREG. The slope was calculated from the DTM using the Horn algorithm (Horn (1981)). The Austrian Federal Forests provided all data.

3.3.3 Reference data

Reference data collection in the field was performed by forest engineers from May 2017 until November 2017 and subsequently digitalised by April 2018. This data acquisition was made as part of the forest management plan update for this region by the Austrian Federal Forests. In particular, forest engineers assess every forest stand by measuring the basal area and tree heights to derive the most important key figures such as tree species composition, yield class and stocking volume. Another part of the elicitation is the planning of thinnings that need to be executed to ensure the optimal growth of the trees and maintain healthy forest stands with high-quality wood. Thinnings are planned in three urgency levels:

- urgency 1 - forest stand needs thinning during the next 0 to 3 years
- urgency 2 - forest stand needs thinning during the next 3 to 10 years
- urgency 3 - forest stand can be thinned at the end of the decennium or can be postponed until the next management plan

The acquired data was populated into a Database, and the spatial information of the forest stands was drawn in a GIS. Subsequently, both data sources were synchronised.

Table 3.2: Reference data classes, their definitions and occupied area.

Class	**Definition**	**Area [ha]**	**Area [%]**
thinning 1	Forest, thinning within 0-3 years	958.95	9.3
thinning 2	Forest, thinning within 3-10 years	1404.88	13.6
thinning 3	Forest, thinning within 9-15 years		
no thinning	Forest, no thinning	7245.62	70.3
other	Non forest (buildings, roads, water bodies)	696.19	6.8

Chapter 4

Methods

Our objective in this study is to predict where and when thinnings are necessary. For this problem statement, we employ deep convolutional neural networks (DCNNs) to generate classifications based on pixels, also known as the semantic segmentation task. This chapter provides information about the data preprocessing, outlines the experimental design we designed to attain our research objectives, describes the training procedure and presents the evaluation criteria applied in this study.

4.1 Data preprocessing

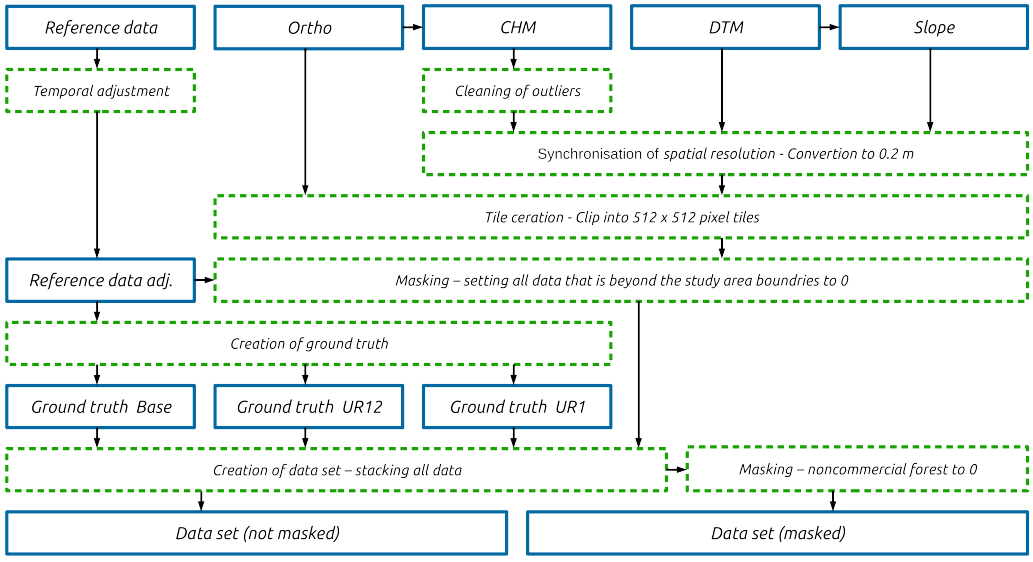


Figure 4.1: Data preprocessing workflow with all data manipulation steps (see text).

Due to different spatial and temporal resolutions (Table 3.1), the described data needed preprocessing in order to apply it as input for the DCNN. The workflow is shown in Figure 4.1 and described in this section in more detail.

4.1.1 Cleaning of outliers

The provided canopy height model (CHM) contained values ranging from -73.9 to 138. Since values below 0 are physically impossible and tree heights above 45 m are improbable in Austria, these values need adjustment. Therefore, values below 0 and over 40 m were highlighted, visually inspected and classified in GIS. The analysis showed that highlighted pixels higher than 47 m and negative values were predominantly rocky steep slopes. Consequently, all pixels with a value beneath 0 and above 47 m were set to 0.

4.1.2 Synchronisation of spatial resolution

The next step in preparing the data was to synchronise the spatial resolution for all input data. As shown in Table 3.1, the orthophotos have a spatial resolution of 0.2 m as opposed to all other raster data having a spatial resolution of 1 m. There are benefits and drawbacks to using either as the standard spatial resolution. Using 0.2 m as the standard results in more detail in the four orthophoto layers. Greater detail can be beneficial for the deep learning algorithm to recognise individual tree crowns and thus help it determine the density of the forest. In contrast, the resolution of 1 m would provide a broader field of view when maintaining the same input image size to the DCNN. We decided to use the resolution of 0.2 m as the standard for all input data as we anticipated the information loss would be too high when using the 1 m resolution. Consequently, all data with a spatial resolution of 1 m was converted using the GDAL library.

4.1.3 Tile size

Furthermore, the data need to be clipped into square tiles as it is the input type of the neural networks used in this study. Accordingly, an image size of 512x512 pixels seems a favourable choice due to the possibility of processing such size with modern GPUs with a decent batch size while maintaining a large field of view. Ultimately the image size could be easily reduced by quartering the dataset without considerable

effort, thereby increasing the batch size accordingly. The image size of 512x512 pixels corresponds to 102.5m x 102.5m in reality. Since one adult tree crown (Norway spruce) has a diameter of 5-6 m, 300-400 adult tree crowns can be represented on one tile.

4.1.4 Tile creation

Considering that only commercial forest is relevant for our study since solely thinnings in a commercially used forest are regularly planned and executed as described in Subsection 3.2, we consequently laid a regular grid of 102.5m x 102.5m over the entire study area to create polygons that we intersected with the reference data. Only polygons that contained over 25% commercial forest area were selected, while the remaining polygons were removed. Finally, we utilised the selected polygons to clip all input raster data into tiles of 512 x 512 pixels.

4.1.5 Reference data adjustment

Considering that the aerial imagery acquisition was recorded in September 2018 and the reference data in May to November 2017, the provided reference data needed to be adjusted. Between the two points in time lies a whole year where cuttings in the study area were executed. We resolved this inconsistency by using data from the Austrian Federal Forests database, where all cuttings are registered, to identify already carried out thinnings and cleaned calamities. Additionally, all commercial forest stands were visually inspected in GIS, and any noticeable errors were corrected.

4.1.6 Masking

Since the reference data is restricted to the study area and we use square tiles, part of the tile frequently contains no information about the reference data (ground truth). Having these parts of the tile with no ground truth information while orthophoto and DTM data is available would be misleading for the machine learning algorithm. Hence, we had to remove any data that was outside the boundaries of the study area. To accomplish that, we created a binary mask from the reference data for every tile and multiplied it with the tiles. In the resulting tiles, all information beyond the boundaries of the study area was set to 0.

Table 4.1: Allocation of the reference data to ground truths *Base*, *UR12* and *UR1* as well as the proportions of pixels representing each class. UR is an abbreviation for urgency. The definitions of the reference data classes are explained in Table 3.2

Reference data	**Base**		**UR12**		**UR1**	
Class	Class	[%]	Class	[%]	Class	[%]
void	void	8.5	void	8.5	void	8.5
thinning 1	thinning	21.0	thinning ur1	8.5	thinning ur1	8.5
thinning 2			thinning ur2	12.5	merged	76.8
thinning 3						
no thinning	no thinning	64.3	no thinning	64.3		
other	other	6.2	other	6.2	other	6.2

4.1.7 Creation of ground truth

Since the creation of the ground truth is firmly connected to our experimental design, we describe both parts of the experimental design and the creation of the ground truth in this subsection. In Table 3.2, we can see the classes of the reference data, their definitions, and the representation in the data set and Figure 4.1 shows the main steps to create the data set.

The first objective we defined in the Section 1 is stating if it is feasible to detect forests in need of thinning from remote sensing data. In order to answer this question, it is simply necessary to distinguish between a forest that needs to be thinned and everything else. When looking at Table 4.1, we can see six unique data classes. Thereby ”void” represents simply the absence of information as described in Subsection 4.1.6. The other five classes are defined in Subsection 3.3.3, of these three are related to thinning. Since we are not interested in the acuteness of the thinnings for this objective, we summarised *thinning 1*, *thinning 2*, and *thinning 3* into one class called *thinning*. This ground truth is called *Base* and it differentiates between the classes *void*, *thinning*, *no thinning* and *other* (Table 4.1).

To address the second objective that states if the urgency of thinnings can be assessed, we created two types of ground truths named *UR12* and *UR1* (Table 4.1). The ground truth labelled *UR12* was designed to differentiate between urgent thinning (*thinning ur1*), not urgent thinning (*thinning ur2*), *no thinning*, *other*, and *void*. Hence, *thinning ur1* was set equivalent with class *thinning 1* of the reference data,

Table 4.2: Allocation of the reference data to masked ground truths *masked Base*, *masked UR12* and *masked UR1* as well as the proportions of pixels representing each class. UR is an abbreviation for urgency. The definitions of the reference data classes are explained in Table 3.2

Reference data	**masked Base**		**masked UR12**		**masked UR1**	
Class	Class	[%]	Class	[%]	Class	[%]
void	void	44.2	void	44.2	void	44.2
thinning 1	thinning	20.4	thinning ur1	8.4	thinning ur1	8.4
thinning 2			thinning ur2	12.0	merged	47.4
thinning 3						
no thinning	no thinning	35.4	no thinning	35.4		

along with *thinning ur2* being defined as *thinning 2* and *thinning 3* grouped together. Subsequently, another ground truth (*UR1*) was created to simplify the problem, thus trying to enhance the model's performance. This ground truth is similar to ground truth *UR12*, with the only difference that the classes *thinning ur2* and *no thinning* are merged into one class named *merged*. As a result, the model can concentrate on finding urgent thinnings by reducing from 5 classes to 4 classes.

Furthermore, we adopt this strategy of addressing the deterioration of performance by reducing the complexity of the problem in the second set of ground truths (*masked Base*, *masked UR12*, *masked UR1*). This second collection of ground truths was created exactly like the first one, except that instead of masking everything beyond the boundaries of the study area as described in subsection masking, we masked everything outside the boundaries of the commercial forest area. The resulting distribution of the classes is presented in Table 4.2. In contrast to the first set of ground truths (*Base*, *UR12*, *UR1*), the masked ground truths (*masked Base*, *masked UR12*, *masked UR1*) contain no class *other*, and there has been a substantial shift in the distribution of the classes towards the class *void*.

Finally, all ground truth data was created by rasterising the reference data (vector data) with the GDAL library into images of size 512x512x1.

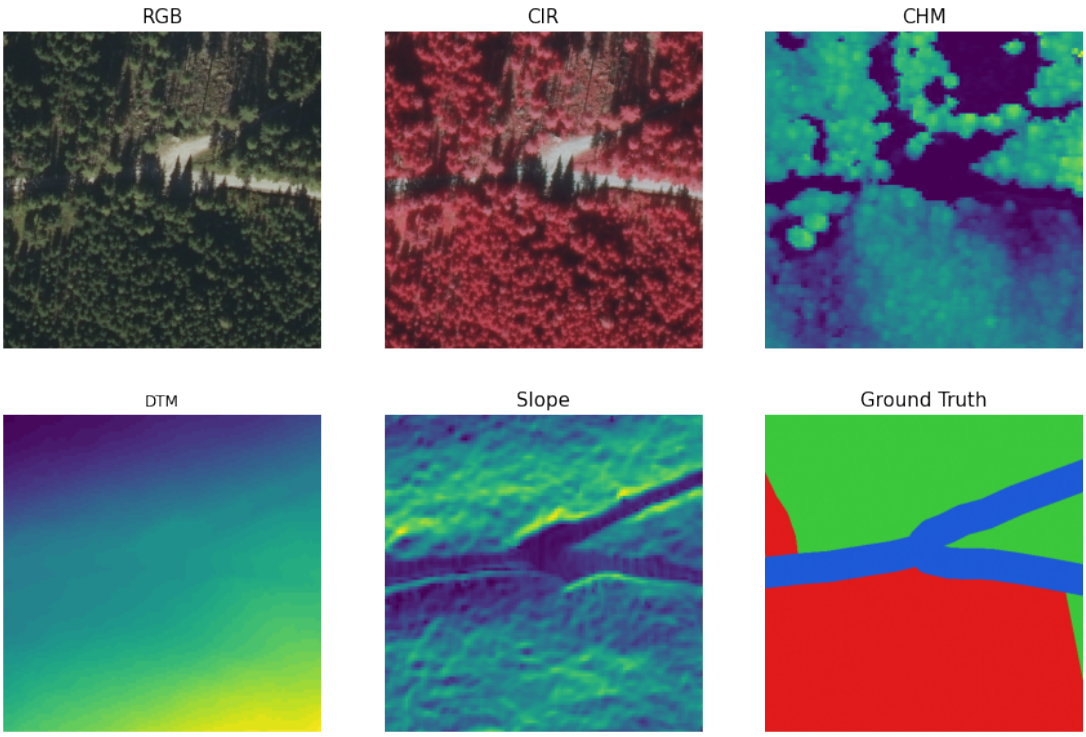


Figure 4.2: Input data tile as used in the final data set (non masked). RGB: true colour orthophoto, CIR: the colour infrared orthophoto, CHM: crown height model, DTM: digital train model, Slope: slope and Ground Truth: ground truth *Base*. Ground Truth classes are coloured where green represents forest not to be thinned (class: forest), red represents forest thinning is necessary (class: thinning) and blue represents everything else (class: other)

4.1.8 Creation of data set

All the preprocessed input data and the generated ground truths were stacked into two data sets. Two separate data sets were needed since the input data corresponding to ground truths *Base*, *UR12* and *UR1* was masked different to the input data of ground truths *masked Base*, *masked UR12*, and *masked UR1*. Hence, we created the data sets by stacking all preprocessed data into two hdf5 files. Each file consists of all tiles created inside the study area, as shown in Figure 3.1. Every tile is composed of the orthophoto (RGB+NIR, 4 layers), the CHM (1 layer), the DTM (1 layer), the Slope (1 layer) and the Ground truths (3 layers), thus summing up to 10 layers (Figure 4.2). Thus, in total per data set, 10250 tiles were created, each of size 512x512x10 pixels resulting in an array with the dimensions 10250x10x512x512.

Subsequently, two more data sets were generated holding the same data, only that

every tile was quartered. Hence, for every 512x512 tile, we produced four 256x256 tiles, which resulted in two data sets with the dimensions 40250x10x256x256. These data sets were generated to accelerate the training process since the smaller tiles can be trained with larger batch size and help the Deep Net to converge quicker.

4.2 Experimental design

Our experimental design can be divided into three major parts and is illustrated in Figure 4.3. In the first part called *Architecture Selection*, we perform a search for the best performing DCNN-architecture. We then analyse the results of the models predicting the necessity of thinning (*Thinning Necessity*) and the urgency of thinnings (*Thinning Urgency*). Finally, we perform an ablation study (*Ablation Study*) to determine the impact of the various input data sources.

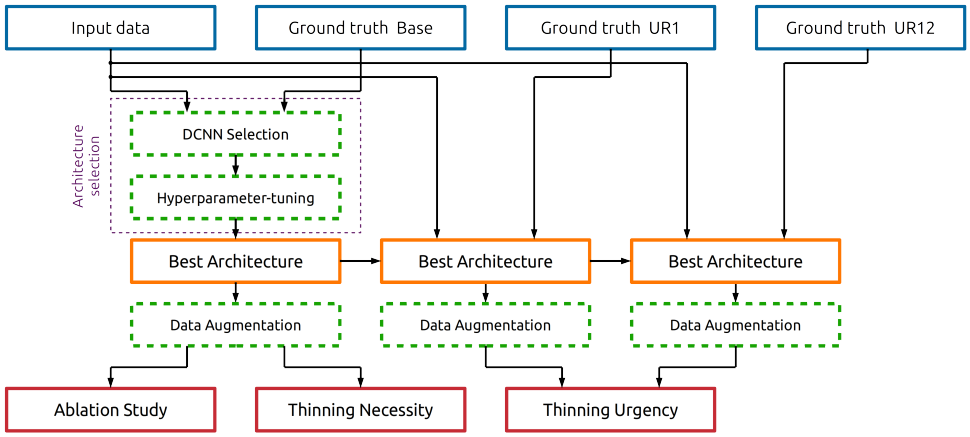


Figure 4.3: Model training workflow with all processing steps to obtain the final models. First stage is the Architecture selection to find the best performing DCNN-architecture. The second stage is the training of the final models to answer the research objectives (*Thinning Necessity*, *Thinning Urgency*, *Ablation Study*).

4.2.1 Architecture selection

The exploration to find the best performing DCNN-architecture was conducted exclusively on the data set with the ground truth *Base* (Table 4.1). The other ground truth types, *UR12* and *UR1*, are later used for clarifying if deep nets can determine thinning urgency. Nonetheless, they are not employed in the search for the best model. From

our literature research outlined in section 2.3, we found the following three DCNNs as the most promising architectures for solving our semantic segmentation problem.

- UNet (Ronneberger et al. (2015))
- FC-DenseNet (Jegou et al. (2017))
- DeepLabv3+ (Chen et al. (2018b))

The exact architectures utilised in this study are illustrated in Appendix C. All DCNNs were modified to accept the tiles from the data set with the dimensions 512 x 512 x 7 (or 256 x 265 x 7) as input and output the predictions as a 512 x 512 x 1 (or 256 x 265 x 1) array.

We search for the optimal architecture by evaluating the three DCNNs defined in n Appendix C on the data set (*Base*). The DCNN achieving the best result is then chosen for further optimisation experiments. This optimisation consists of manipulating individual parts of the architecture to further optimise the model's performance by applying the Bayesian optimisation method (Snoek et al. (2012)). This method is an alternative to an exhaustive full grid-search, which is computationally very expensive. Instead, it efficiently searches for the optimal hyper-parameters based on the Bayes Theorem. In practice, that means we sequentially alter critical structures in the DCNN-architecture and always adopt the structure that provides the best results.

All DCNNs are implemented in PyTorch and all supplementary code was written in Python. All code is freely available at `https://github.com/satlawa/edin_thinning_necessity`. The experiments were performed on a system with Ubuntu 20.04 as the operating system equipped with an I7 6700K CPU, 32GB of RAM and an Nvidia RTX 3090.

4.2.2 Thinning Necessity and Urgency

After finding the best architecture, we train the final models on the two data sets with all six ground truths (*Base*, *UR12*, *UR1*, *masked Base*, *masked UR12*, *masked UR1*). These final models are then evaluated on the test set to estimate the actual performance on unseen data. Subsequently, we analyse the results and determine whether the models can satisfactorily fulfil the intended task. In the case of the ground truth *Base* and *masked Base*, we examine if the necessity of thinnings can be detected by a DCNN using solely remote sensing data. Furthermore, we employ the ground truths *UR12*,

masked UR12, *UR1*, and *masked UR1* to investigate if the urgency of thinnings can be determined in addition to detecting the necessity of thinnings.

4.2.3 Ablation study

Finally, we perform an ablation study by omitting different types of input data based on our data set with ground truth *Base*. The primary purpose of this experiment is to determine the importance of the individual input data types as shown in Table 3.1.

4.3 Training

Before training, we randomly shuffle the data set and use 70% of the data for training, 10% for validation and the remaining 20% for testing. Accordingly, we employ this split for training and evaluating all models. Thus, the test set is exclusively applied to the best performing model chosen by evaluation on the validation set. Furthermore, all input data were standardised by subtracting the mean and dividing by the standard deviation for each value of each input channel (3.1).

All models are trained from scratch due to the dissimilarity of the input data compared to the data sets used on pre-trained models. For initialisation of the weights, we employ the Kaiming uniform initialisation (He et al. (2015a)). For all our experiments, we used the Adam optimiser as in Kingma and Ba (2015), with an initial learning rate for (1) UNet of 0.01, (2) FC-DenseNet of 0.003 and (3) DeepLabv3+ of 0.001. After every epoch, we applied a decay rate of 0.995 to the learning rate. Since loss functions play a decisive role in training models, choosing a suitable loss function is essential (Jadon (2020)). In our case, the distribution of the classes is skewed. Therefore we implement and apply the dice loss as our loss function. The dice loss is defined as 1 - F_1 score (defined in Equ. 4.5). The batch size was chosen to be as large as possible, with the constraint being the memory of the GPU. Depending on the DCNN architecture, the batch size was ranging between 16 and 72. Due to long training times (up to 40 minutes for one epoch), only one 5 fold cross-validation on the model *Base* was carried out. The models were trained until convergence, however at least for 50 epochs. Data augmentation in the form of horizontal flips was performed only on the previously chosen best-performing architecture. We restricted the data augmentation to horizontal flips due to the different growing conditions on north and south slopes.

4.4 Evaluation

To evaluate the performance of the trained models thoroughly, we use five evaluation metrics. The first criterion is the overall accuracy (Acc) which is determined by dividing the number of all correctly classified pixels by the total number of pixels (Equation 4.1).

$$Acc = \frac{\sum_{i=1}^{n}(TP_i + TN_i)}{\sum_{i=1}^{n} p_i} \tag{4.1}$$

Where n is the number of images; TP_i is the number of true positive pixels in image i; p_i is the number of pixels in image i.

Although Acc is a very intuitive and thus common evaluation criteria, it is not very meaningful in cases where the distribution of classes is highly skewed, as is the case with the data set we utilise (Table 3.2). Therefore, we apply the metrics precision (Equation 4.3), recall (Equation 4.2), IoU (Equation 4.4) and F_1 score (Equation 4.5).

$$recall_j = \frac{\sum_{i=1}^{n} TP_{ij}}{\sum_{i=1}^{n}(TP_{ij} + FN_{ij})} \tag{4.2}$$

$$precision_j = \frac{\sum_{i=1}^{n} TP_{ij}}{\sum_{i=1}^{n}(TP_{ij} + FP_{ij})} \tag{4.3}$$

$$IoU_j = \frac{\sum_{i=1}^{n} TP_{ij}}{\sum_{i=1}^{n}(TP_{ij} + FP_{ij} + FN_{ij})} \tag{4.4}$$

$$F_1 - Score_j = 2 * \frac{precision_j * recall_j}{precision_j + recall_j} \tag{4.5}$$

Where n is the number of images; TP_{ij} is the number of pixels in image i, which are correctly predicted as class j; FP_{ij} is the number of pixels in image i, which are incorrectly predicted as class j; FN_{ij} is the number of pixels in image i, which are incorrectly predicted as any class other than class j.

Since we have a multi-class problem statement, we first calculate the proposed metrics per class and determine the mean among all classes as illustrated in Equation 4.6.

$$mX = \frac{1}{m} \sum_{j=1}^{m} X_j \tag{4.6}$$

Where m is the number of classes; X_j is one of the metrics (precision, recall, IoU, F_1) for class j.

Despite providing all the above metrics for a holistic evaluation of the models, we adopt the F_1 score as the sole decisive evaluation score. Even though the ground truth contains the class void we omit this class from the calculation of all metrics since it is of no use for our problem. Furthermore, the void class is very easy to learn for the classifier, hence its presence in the metrics would distort the results and indicate a better model performance than the particular model can achieve in reality. All evaluation on the test set was performed on the 512x512 data set.

Despite providing all the above metrics for a holistic evaluation of the models, we adopt the F_1 score as the sole decisive evaluation score. Furthermore, class *void* carries essentially no information. However, it is still necessary as the data outside the study area boundaries of the commercial forest boundaries (in the case of masked ground truths) must be represented. Nonetheless, although we require class void for training the DCNNs, we omit it in the presentation of the results as it was nearly perfectly classified and carried no valuable information. Besides, its presence in the metrics would distort the results and indicate a better model performance than the particular model can achieve in reality. Finally, all evaluation on the test set was performed on the 512x512 data set.

Chapter 5

Results

This chapter outlines and interprets the results of the performed experiments based on the previously described methods in Chapter 4. In particular, we present the selection of the best performing model. In addition, we demonstrate the feasibility of using DCNN for classifying the urgency of thinnings as well as the ablation study.

5.1 Architecture Selection

5.1.1 DCNN Selection

We start by comparing the performance of the three DCNN architectures UNet, FC-DenseNet and DeepLabv3+. Subsequently, we select the architecture based on its achievement on the ground truth *Base*. We can see in Table 5.1 DeepLabv3+ performs best with an F_1 score of 80.38%. It seems that the atrous convolutions of DeepLabv3+ capture the best contextual information. Hence, we chose DeepLabv3+ as the default network architecture and perform the hyper-parameter tuning on this DCNN.

Table 5.1: Performance of three selected DCNN-architectures on the validation set of ground truth *Base*. Detailed information about the architectures is provided in Appendix C. Acc: overall accuracy, mIoU: mean intersection over union and F1: F_1 score

DCNN	Acc	Precision	Recall	mIoU	F1
UNet	85.78%	82.39%	78.11%	67.05%	79.82%
FC-DenseNet	80.03%	76.33%	70.82%	57.91%	72.62%
DeepLabv3+	85.83%	81.88%	79.47%	67.76%	**80.38%**

5.1.2 Hyper-parameter tuning

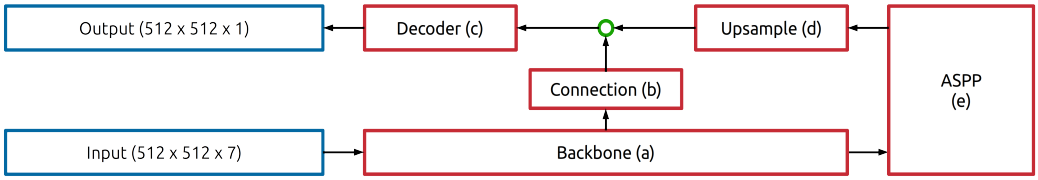


Figure 5.1: Modification parts of DeepLabv3+. Parts represent by letters.

After determining the best performing DCNN trained with the standard hyper-parameters, we optimise the hyper-parameters of DeepLabv3+. Although the DeepLabv3+ architecture is already tuned on the PASCAL VOC 2012 data set, the data set we use is considerably divergent from the PASCAL VOC 2012. That is why we optimise the hyper-parameters by applying the Bayesian optimisation method. In particular, we examine the performance of the DCNN by modifying five critical parts of the DeepLabv3+. Figure 5.1 illustrates the experimental design employed to obtain the best model.

Table 5.2: Effect of backbone (part a) on the validation set performance (ground truth *Base*). Acc: overall accuracy, mIoU: mean intersection over union and F1: F_1 score

Backbone	**Acc**	**Precision**	**Recall**	**mIoU**	**F1**
Xception	85.83%	81.88%	79.47%	67.76%	80.38%
ResNet 101	86.81%	82.94%	79.85%	69.00%	**81.26%**

We begin the optimisation process by exchanging the modified Xception architecture for the Resnet 101 architecture as the backbone module of the DCNN (part a). This module is responsible for encoding the features from the initial images until they are passed to the ASPP module. As Table 5.2 shows, replacing the backbone module to the Resnet 101 architecture results in an F_1 score of 81.26%, which is a 0.88% gain compared to the 80.38% achieved with the Xception architecture. Hence, we fix Resnet 101 as the backbone for our DeepLabv3+ net for all further experiments.

We further alter the [1 x 1, 48] convolution (part b), responsible for providing the low-level feature map information of the encoder to the decoder. Correspondingly, we modify the number of channels as shown in Table 5.3. From the results in Table 5.3, we deduct that neither increasing nor decreasing the number of filters is helping to

Table 5.3: Effect of the convolution connecting the encoder to the decoder (part b) on the validation set performance (ground truth *Base*). Acc: overall accuracy, mIoU: mean intersection over union and F1: F_1 score

Connection	**Acc**	**Precision**	**Recall**	**mIoU**	**F1**
[1 x 1, 32]	86.77%	83.71%	79.19%	68.80%	81.11%
[1 x 1, 48]	86.81%	82.94%	79.85%	69.00%	**81.26%**
[1 x 1, 64]	86.53%	82.04%	80.53%	68.73%	81.07%

improve the scores. Thus we retain the initial 48 filters.

Subsequently, we experiment with the upsampling convolution (part c), which is accountable for upsampling the encoded features from the ASPP module to the decoder by doubling the number of filters from 256 to 512 for successive convolutions. We report the findings in Table 5.4 and show that increasing the number of channels degrades performance. Hence, we leave the network architecture with its original [1 × 1, 256] x 2 upsampling convolution.

To evaluate the effect of the decoder (part d), we perform manipulations on the number of channels and the number of convolutions. All variations of the decoder and their performance are reported in Table 5.5. As a result, we can conclude that applying convolutions with 256 channels provides the best results, whereas adding or removing channels results in performance deterioration. Similarly, the experiment of adding additional convolutions brings no benefit and results in performance loss. Hence, we leave the initial [3 × 3, 256] x 2 convolutions combination unchanged.

Lastly, we alter the output stride (part e), defined as the ratio between input image spatial resolution and final output resolution, and perform data augmentation. By replacing the output stride 16 to 8, we remove one more block from the ResNet 101,

Table 5.4: Effect of the upsampling convolution (part c) on the validation set performance (ground truth *Base*). Acc: overall accuracy, mIoU: mean intersection over union and F1: F_1 score

Up Sample	**Acc**	**Precision**	**Recall**	**mIoU**	**F1**
[1 × 1, 256] x 2	86.81%	82.94%	79.85%	69.00%	**81.26%**
[1 × 1, 512] x 2	86.62%	82.64%	79.57%	68.63%	80.99%

Table 5.5: Effect of the decoder (part d) on the validation set performance (ground truth *Base*). Acc: overall accuracy, mIoU: mean intersection over union and F1: F_1 score

Decoder	**Acc**	**Precision**	**Recall**	**mIoU**	**F1**
[3 × 3, 128] x 2	86.55%	68.83%	82.50%	79.87%	80.94%
[3 × 3, 256] x 2	86.81%	69.00%	82.94%	79.85%	**81.26%**
[3 × 3, 256] x 3	86.74%	68.83%	82.80%	80.00%	81.13%
[3 × 3, 256] x 4	86.67%	68.44%	82.87%	79.24%	80.83%
[3 × 3, 512] x 2	86.60%	68.89%	82.87%	80.09%	81.19%
[3 × 3, 512] x 3	86.45%	68.73%	81.96%	80.46%	81.08%
[3 × 3, 1024] x 2	86.60%	68.49%	82.89%	79.19%	80.88%

employ atrous convolution with rate 4 instead, and gain a denser feature extraction. As a result, applying output stride 8 helps achieve a better outcome of 81.59% on the F_1 score (Table 5.6). Furthermore, by employing horizontal flipping on the input data, we double the training data size and increase performance to 83.01%. Hence using stride 8 and applying data augmentation outperforms all other combinations.

Finally, we adopt the best performing hyper-parameters and employ this DCNN-architecture to train all subsequent models. A detailed diagram of the final architecture is provided in Appendix B.

Table 5.6: Effect of the output stride (part e) and data augmentation on the validation set performance (ground truth *Base*). Acc: overall accuracy, mIoU: mean intersection over union, F1: F_1 score, OS: output stride, Flip: Adding horizontally flipped inputs.

Os	**Flip**	**Acc**	**Precision**	**Recall**	**mIoU**	**F1**
8	no	86.84%	82.77%	80.77%	69.43%	81.59%
16	no	86.81%	82.94%	79.85%	69.00%	81.26%
8	yes	88.17%	85.17%	81.20%	71.45%	**83.01%**
16	yes	87.69%	83.76%	82.01%	71.03%	82.74%

5.2 Thinning

After optimising the DCNN-architecture on the data set, we focus on the main research objectives of this study. First, we evaluate the possibility of detecting the need for thinning with the optimised DCNN exclusively with remote sensing data (subsection 5.2.1). Thereupon, we respecify the research objective to detect the need of thinning with urgency and assess its feasibility (subsection 5.2.2).

5.2.1 Thinning necessity

Table 5.7: Class scores and mean class scores on the test set (ground truth *Base*) with 5-fold cross validation. The model's objective is to detect the necessity of thinning. Class definitions, *thinning*: thinning within 1-10 years, *no thinning*: no thinning necessary, *other*: non forest areas. std: standard deviation.

Score	thinning	no thinning	other	mean	std
Precision	77.08%	91.58%	79.69%	82.78%	0.78%
Recall	80.32%	90.92%	74.15%	81.79%	0.52%
IoU	64.81%	83.89%	62.34%	70.35%	0.44%
F1	78.64%	91.24%	76.80%	**82.23%**	**0.31%**

Based on the network architecture from 5.1.2, we evaluate the model on the *Base* test set (Table 5.7). With a mean F_1 score of 82.23%, the model achieves a similar score to the 83.01% on the *Base* validation set. Thus, we can conclude that the chosen model is not overfitting to the validation set. Additionally, when focusing on the class-specific scores, we see the model performing best on predicting the class *no thinning*, whereas the scores of the other two classes are significantly lower.

Consequently, when examining the confusion matrix in Table D.3, we gain further insight into the misclassifications of the model. Accordingly, we identify that the main mistakes happen between the classes *thinning* and *no thinning* as well as between *no thinning* and *other*, while misclassifications between the classes *thinning* and *other* are insignificant. These results match our knowledge about the classes. Examples of predictions are illustrated in Figure 5.2.

For instance, class *thinning* represents dense forest with a minimum top height of around 13 m and thus contrasts to class *other* that often represents no vegetation or

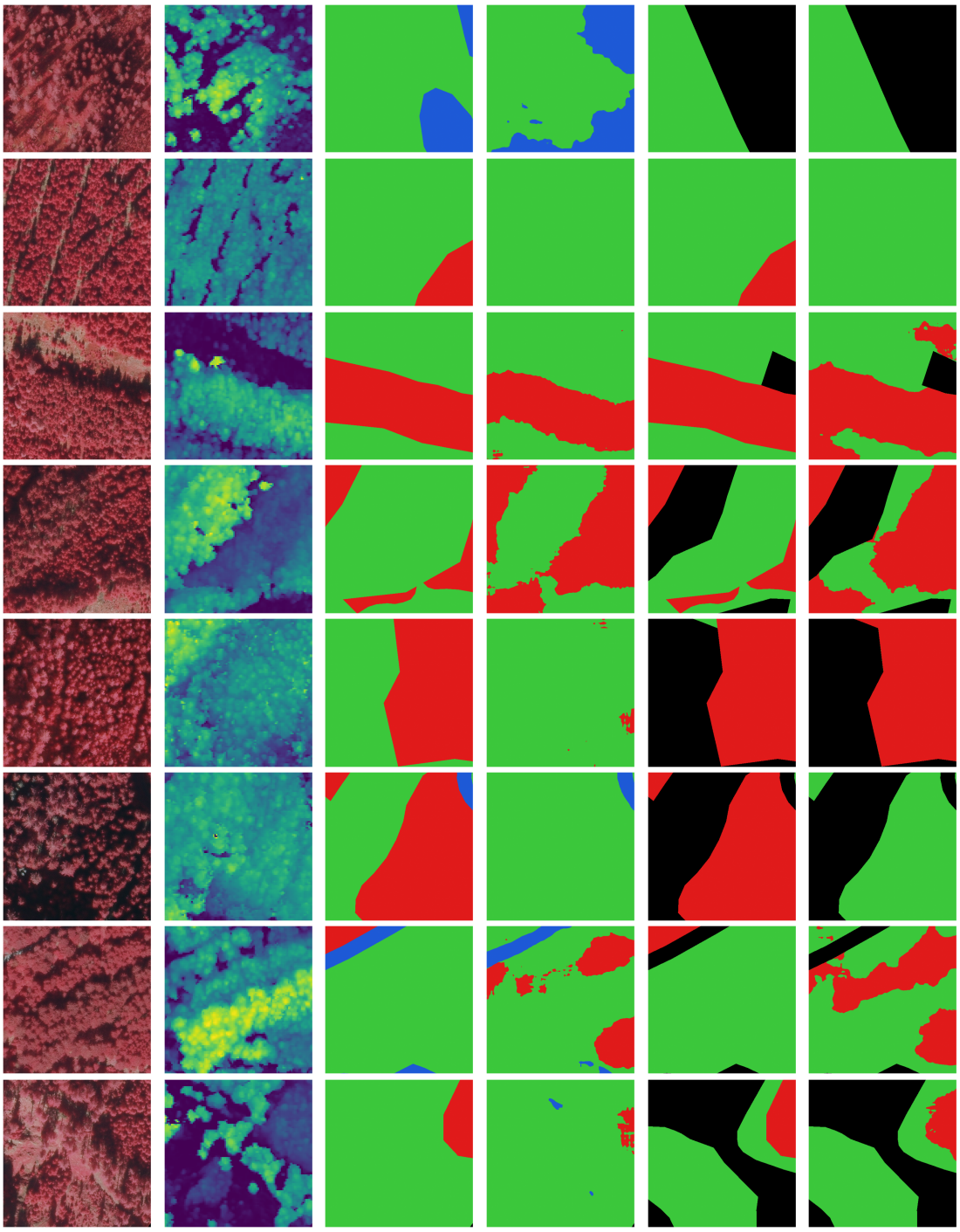
Table 5.8: Confusion matrix on test set (ground truth *Base*). The numbers represent classified pixels in 10^6.

		prediction			
	Class	**thinning**	**no thinning**	**other**	∑
reference	**thinning**	89	26	1	116
	no thinning	22	341	5	368
	other	1	7	24	32
	∑	112	374	30	516

shallow growing vegetation like grassland or mountain pines. In contrast, it makes sense to see the model misclassifying the class *no thinning* with both other classes as it embodies forest in all ages. For example, a very young forest has almost identical features compared to grassland, as presented in Figure 5.2 row 1. Therefore, it is challenging and sometimes impossible to distinguish between *no thinning* and *other* with only remote sensing data. Equally ambitious are some classification cases between the classes *thinning* and *no thinning*. In this situation, the model struggles to differentiate between edge cases of dense forest (Figure 5.2 rows 5 to 8). These errors might arise due to lacking information about the yield class of the forest. For instance, dense old forest on less vigorous sites might appear similar to a dense middle-aged forest on productive sites or vice versa (Figure 5.2 row 6). Likewise, young forest on very viable sites might already require thinning during the planning period whereas similar forest on less viable sites grows slower and should therefore not be planned for thinning.

Unquestionably, the model produces misclassifications, yet the results are excellent for most cases. For example, it correctly classifies the recently thinned forest as not to be thinned (Figure 5.2 row 2). Similarly, the model has no problems classifying correctly the cut forest and the more sparsely standing forest as *no thinning*, whereas the dense forest it accurately predicts as *thinning* (Figure 5.2 row 3 and 4). Finally, the model was used on the entire study area to create a final map that we provide in Appendix A.

As we are particularly interested in the differentiation between forest with and without the necessity of thinning, we trained another model that discriminates just among the classes *thinning* and *no thinning*. For accomplishing this, we employed the data set containing the ground truth *masked Base*, which contains just information about commercial forest while all other data was masked.



(a) CIR (b) CHM (c) GT (d) Pred (e) GT-m (f) Pred-m

Figure 5.2: Examples of semantic segmentation on test set. Column (a) CIR are false colour composites with near infrared, (b) CHM is the canopy height model, (c) GT is the ground truth *Base*, (d) is the prediction of the model trained on the ground truth *Base*, (e) GT-m is the ground truth *masked Base*, (f) is the prediction of the model trained on the ground truth *masked Base*. In (b) the colour palette illustrates low heights as dark (dark blue) and high heights as bright (yellow). The colours in (c) until (f) represent the following classes, black: void, red: thinning, green: no thinning, blue: other.

Table 5.9: Class scores and mean class scores on the test set (ground truth *masked Base*).The model's objective is to detect the necessity of thinning restricted to the commercial forest.

Score	**thinning**	**no thinning**	**mean**
precision	76.98%	92.27%	84.63%
recall	85.84%	86.82%	86.33%
IoU	68.31%	80.93%	74.62%
F1	81.17%	89.46%	**85.32%**

As reported in Table D.8 we obtain a mean F_1 score of 85.32%. Hence, by focusing on merely two classes, we raised the mean F_1 score by 2.85% compared to the data set without masking (Table 5.7). When concentrating on the class-specific scores, we see that the mean F_1 score's gain is due to the more accurate classification of the class *thinning*.

More detailed information of the misclassifications between the classes *thinning* and *no thinning* is provided in the confusion matrix in Table 5.10. What is striking is the lower number of categorised pixels due to the masking of non-commercial forest areas and the much higher number of false positives compared to the true negatives. This finding contrasts with the first model we evaluated (Table D.3), where the false positives and the true negatives were relatively balanced. Precisely this behavioural distinction can be observed in the examples of Figure 5.2. Particularly apparent is the difference in sensitivity in the fifth example of Figure 5.2, where the prediction of the non-masked data set (d) predicts *no thinning* on the entire tile while the prediction of the masked data set (f) classifies the entire commercial forest as in need of thinning.

Table 5.10: Confusion matrix on test set (ground truth *masked Base*). The numbers represent classified pixels in 10^6.

		prediction		
	Class	**thinning**	**no thinning**	$\sum$
ref.	**thinning**	97	16	113
	no thinning	29	191	220
	$\sum$	126	207	333

5.2.2 Thinning urgency

Table 5.11: Class scores and mean class scores on the test set (ground truth *UR12*). The model's objective is to detect the urgency of thinnings. Class definitions, *thinning ur1*: thinning within 1-3 years, *thinning ur2*: thinning within 3-10 years, *no thinning*: no thinning necessary, *other*: non forest areas.

Score	thinning ur1	thinning ur2	no thinning	other	mean
precision	55.17%	46.15%	92.35%	80.00%	68.42%
recall	32.65%	71.64%	88.59%	75.00%	66.97%
IoU	25.81%	39.02%	82.53%	63.16%	52.63%
F1	41.03%	56.14%	90.43%	77.42%	66.25%

The research question we are addressing here is whether it is possible to detect the need of thinnings and predict their urgency accurately. Compared to the model in subsection 5.2.1, this question adds another layer of complexity to the network since it has to assess the thinnings urgency. For answering this question, we employ ground truths *UR12* and *UR1* from the data set *Base*. The idea of ground truth *UR12* is to predict thinnings and their urgency directly. Hence it differentiates between very urgent thinnings (*thinning ur1*), less urgent thinnings (*thinning ur2*) and no need for thinning *no thinning*. Moreover, the ground truth *UR1* focuses solely on very urgent thinnings, thereby trying to achieve better performance than *UR12*. A detailed description of the ground truths is provided in the section 4.1.7.

Examining the results in Table 5.11, we instantly perceive the drop of the mean F_1

Table 5.12: Confusion matrix on test set (ground truth *UR12*). The numbers represent classified pixels in 10^6.

		prediction				
	Class	**thinning ur1**	**thinning ur2**	**no thinning**	**other**	$\sum$
reference	**thinning ur1**	16	24	9	0	49
	thinning ur2	8	48	11	0	67
	no thinning	5	31	326	6	368
	other	0	1	7	24	32
	$\sum$	29	104	353	30	516

score from 82.23% achieved with the best model trained on *Base* to 66.25% trained on *UR12*. However, when comparing the scores of the classes *no thinning* and *other* in tables Table 5.7 and Table 5.11, we notice almost no change between the models *Base* and *UR12*. Nevertheless, the F_1 scores of classes *thinning ur1* and *thinning ur2* with 41.03% and 56.14%, respectively, are moderate compared with the 78.07% achieved with the *Base* model on *thinning*.

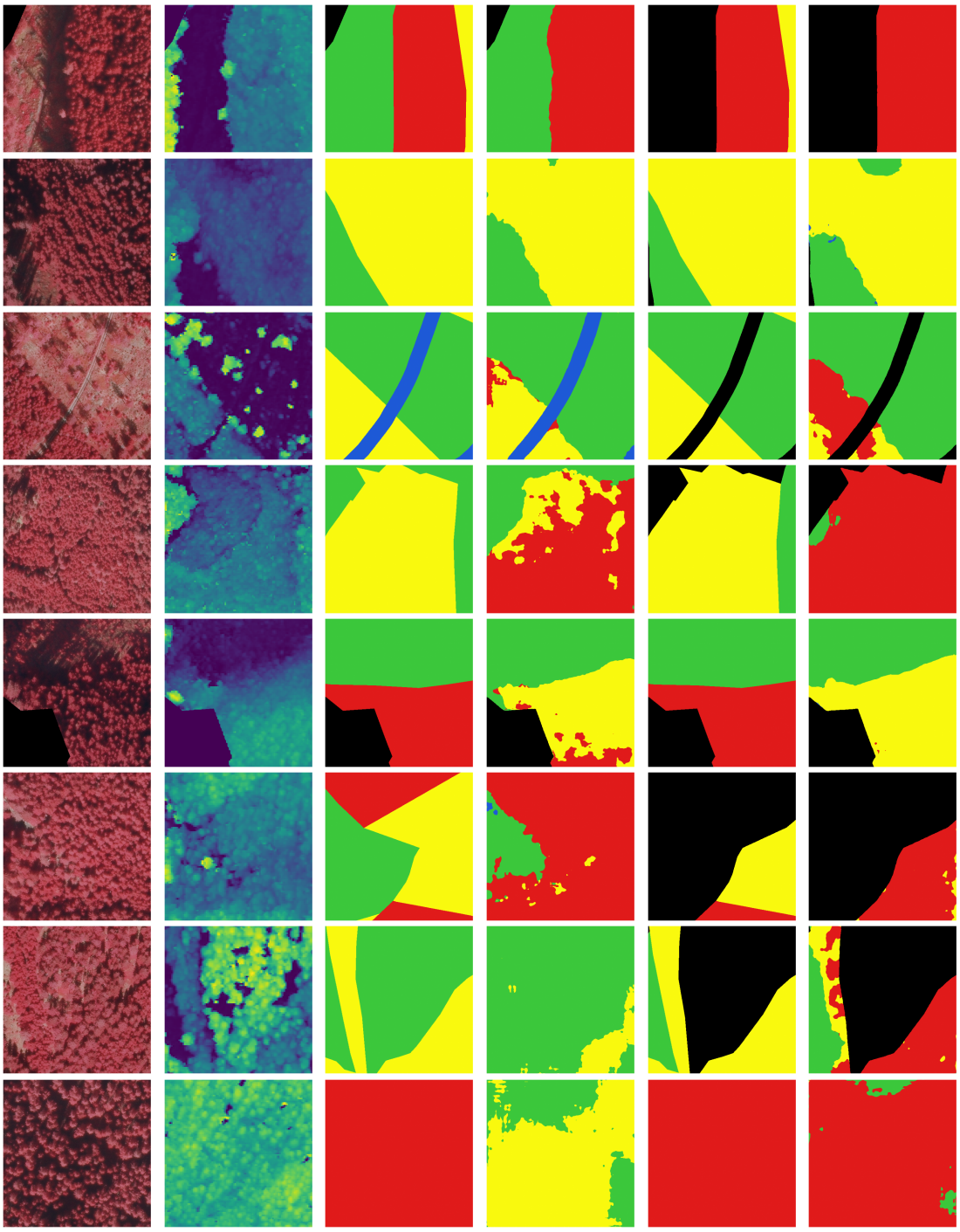
The confusion matrix in Table 5.12 substantiates that the model struggles mainly between the classes *thinning ur1* and *thinning ur2*. Consequently, when we examine the examples in Figure 5.3, we recognise that the model predicts the area with the necessity of thinning rather well. However, although the model produces partially excellent classifications (Figure 5.3 row 1 to 3), it frequently has difficulties classifying the urgency of the thinning correctly (Figure 5.3 row 4 to 8) as anticipated from the confusion matrix.

Accordingly, we can attribute this deterioration fully to the more arduous task of identifying the urgency of thinnings. This finding coincides with the fact that assigning the urgency of thinning is also a difficult task for the experts in the field and thus usually holds a subjective component. Furthermore, we can interpret the almost constant performance on the classes *other* and *no thinning* on the models *Base* and *UR12* as a sign that the DCNN learned in both cases similar features and thus provides us with confidence that the network is well-tuned.

When constraining the problem by just taking into account commercial forest, we were able to increase the performance slightly in the case of ground truth *masked Base* (subsection 5.2.1). Accordingly, we applied the same strategy on the model *masked UR12*. However, the results show no significant overall improvement in the mean F_1

Table 5.13: Class scores and mean class scores on the test set (ground truth *masked UR12*). The model's objective is to detect the urgency of thinnings restricted to the commercial forest.

Score	**thinning ur1**	**thinning ur2**	**no thinning**	**mean**
precision	42.72%	50.32%	92.97%	62.00%
recall	64.74%	48.84%	83.19%	65.59%
IoU	34.66%	32.95%	78.27%	48.63%
F1	51.48%	49.57%	87.81%	62.95%



(a) CIR (b) CHM (c) GT (d) Pred (e) GT-m (f) Pred-m

Figure 5.3: Examples of semantic segmentation on test set. Column (a) CIR are false colour composites with near infrared, (b) CHM is the canopy height model, (c) GT is the ground truth *UR12*, (d) is the prediction of the model trained on the ground truth *UR12*, (e) GT-m is the ground truth *masked UR12*, (f) is the prediction of the model trained on the ground truth *masked UR12*. In (b) the colour palette illustrates low heights as dark (dark blue) and high heights as bright (yellow). The colours in (c) until (f) represent the following classes, black: void, red: thinning ur1, yellow = thinning ur2, green: no thinning, blue: other.

(Table 5.13) over model *UR12*. Whereas the model *masked UR12* achieves a higher F_1 score in class *thinning ur1* (51.48%) compared to model *UR12* (41.03%), all other class scores achieve worse performance. Since we are particularly interested in predicting the urgent thinnings well, model *masked UR12* provides us with a slightly better alternative than model *UR12*. Thus we can conclude that the restricted model *masked UR12* can gain an improvement in class *thinning ur1* compared to the model *UR12*.

Table 5.14: Class scores and mean class scores on the test set (ground truth *UR1*). The model's objective is to detect very urgent thinnings. Class definitions, *thinning ur1*: thinning within 1-3 years, *merged*: thinning within 3-10 years and no thinning needed, *other*: non forest areas.

Score	**thinning ur1**	**merged**	**other**	**mean**
precision	47.29%	93.03%	78.67%	73.00%
recall	54.97%	91.72%	73.98%	73.56%
IoU	34.09%	85.82%	61.62%	60.51%
F1	50.84%	92.37%	76.25%	73.16%

Finally, we reduce the problem even further by focusing on the urgent thinnings (*thinning ur1*) in the ground truths *UR1* and *masked UR1*. As in former cases, the difference between *UR1* and *masked UR1* is the restriction of the latter to only the commercial forest. Correspondingly are the results of *UR1*, displayed in Table 5.14 and the ones of *masked UR1* in Table 5.15.

Table 5.15: Class scores and mean class scores on the test set (ground truth *masked UR1*). The model's objective is to detect very urgent thinnings restricted to the commercial forest.

Score	**thinning ur1**	**merged**	**mean**
precision	44.64%	93.08%	68.86%
recall	61.98%	86.93%	74.46%
IoU	35.05%	81.65%	58.35%
F1	51.90%	89.90%	70.90%

By simplifying the problem hence merging the classes *thinning ur2* and *no thinning*, model *UR1* and *masked UR1* are capable of achieving similar performance on

class *thinning ur1* while eliminating the misclassifications between *thinning ur2* and *no thinning*. Even though models *UR1* and *masked UR1* produce a marginally better performance than *masked UR1*, all models struggle to provide the performance needed for placing them in production. The confusion matrices of models *masked UR12*, *UR1* and *masked UR1* are provided in Appendix D.

5.3 Ablation study

Table 5.16: Ablation study with F_1 class scores and mean F_1 class scores on the *Base* test set. The various input features are abbreviated. O: Orthophoto, C: CHM (Crown height model), D: DTM (Digital train model), S: Slope

Model	Input	thinning	no thinning	other	mean
a	**O+C+D+S**	78.07%	**91.91%**	77.42%	82.47%
b	**O+C+D**	78.32%	91.14%	73.48%	80.98%
c	**O+C+S**	**78.65%**	91.31%	77.78%	**82.58%**
d	**O+D+S**	75.31%	90.69%	**77.96%**	81.32%
e	**C+D+S**	74.12%	89.14%	73.93%	79.07%
f	**O+C**	78.13%	91.38%	76.05%	81.85%
g	**D+S**	45.78%	76.46%	73.03%	65.09%
h	**O**	76.79%	90.84%	75.68%	81.10%

The ablation study investigates the importance of the individual input features by training models with various combinations of input features removed. The performance gains and collapses of the different models provide insight into what effect a specific input feature has on the model. Table 5.16 shows the results of the ablation study with all variants of input feature sets.

From the results, we can deduce that the *DTM* has no or a slightly negative impact on the model's performance. Models trained without *DTM* (*model a* and *model b*) performed slightly better on the mean F_1 score than the models where *DTM* was included (*model c* and *model f*). When examining the class scores for these models, it is apparent that it is primarily in the class *other* where the *DTM* deteriorates the performance. Thus, the input feature *DTM* appears not to contain any useful information that the DCNNS can exploit.

When examining the input feature *Slope*, we can identify a better performance of models that include *Slope* (*model a* and *model c*) compared to the models that exclude *Slope* (*model b* and *model f*). In both comparisons, *model a* versus *model b* and *model c* versus *model f*, the F_1 class score of *other* increases, while all other class scores remain nearly constant. Hence, we can infer *Slope* holds peculiar information about the class *other* that no other input feature contains.

Since *Slope* is calculated from the *DTM*, hence they both share the same data basis, and we expected that the DCNN could at least partially learn to extract some useful information out of the *DTM*. However, it seems the DCNN is not capable of deriving *Slope* out of the *DTM*. Furthermore, when exploring the class scores of model *model g*, we see the model *model g* performing worst in the classes *thinning* and *no thinning* of all the trained models. From this, we can reason that the input features *Slope* and *DTM* hold no information about the forest, just information about the terrain, which conforms with the knowledge we have about these input features. Furthermore, when training just on DTM and slope as input features, we see a good performance in the *other* class as well as a poor performance on the *thinning*, *no thinning* classes. That is per the fact that DTM and CHM contain no information about the forest, just information about the terrain.

Withholding the input feature *CHM* results in models (*model d* and *model h*) having a lower score in the classes *thinning* and *no thinning* compared to the equivalent models with *CHM* included (*model a* and *model f*). Consequently, we can conclude that *CHM* contains some unique information about the forest that the models can exploit to better discriminate between forest with the necessity of thinning and forest that does not have this necessity. Hence, the model can better separate since tree heights are essential criteria for assessing the necessity of thinning, as section 3.2 briefly outlines.

By omitting the input feature *Orthophotos* (*model e*), we can see a significant decline in performance compared to the full model (*model a*). As with *CHM*, the feature *Orthophotos* increases the scores on classes *thinning* and *no thinning*. It moreover improves the score of class *other*. Moreover, when training solely with *Orthophotos* as an input feature (*model h*), we obtain excellent mean F_1 scores, 1.37% worse when compared with the full model (*model a*). Given these results, we can deduce that *Orthophotos* contains most of the valuable information of all input features. The results coincide with the finding that the input feature *Orthophotos* is also the most informative for the forest manager when assessing the forest.

Chapter 6

Conclusion

Although on-time planning and execution of thinnings are crucial for maintaining a healthy forest, minimal research has been performed to derive the need for thinning directly from remote sensing data. Accordingly, we presented in this study the potential of predicting the necessity of thinning with state of the art deep learning architectures solely from very high remote sensing data.

Using multispectral orthophotos, canopy height model (CHM), digital train model (DTM) and slope, and the reference data collected in the field by experts, we created two data sets to answer the research objective. First, we explored three different DCNN-architectures for semantic segmentation whereby the DeepLabv3+ architecture was found to be the best-suited DCNN for the task of detecting the necessity of thinnings. Then, after fine-tuning the DCNN-architecture, we employed the best performing model on the test set, achieving an F_1 score of 82.23% and proving that deep learning algorithms are highly beneficial for detecting forests in need of thinning from remote sensing data. In addition, we were able to increase the performance even further by simplifying the problem and reducing the number of predicted classes from three to two. Finally, we employed the masked data set, which is restricted exclusively to the commercial forest and reached an F_1 score of 85.32%.

From these results, we can deduce that the DCNN was able to retrieve critical information about the density of the forest from the remote sensing data to assess the need for thinning. We can draw this conclusion since the necessity for thinning is determined mainly by two criteria: tree heights and the standing density (basal area) of a forest stand, and the tree heights are already part of the input data. Thus, the model was capable of deriving the crown density and not the actual basal area. Therefore, the crown density is sufficient for the target thinning type, which is crown thinning.

However, for other thinning types, such as low thinning, the DCNN might struggle to produce comparable results. Furthermore, the model would probably provide moderate performance since only suppressed and sub-dominant trees are removed in low thinning, which have no impact on the canopy.

Besides creating a model detecting forests in need of thinning, we additionally tried to predict the urgency of thinnings. Nevertheless, the trained models struggled to distinguish between urgent and not urgent thinning and provided unsatisfactory performance. The poor performance is possibly due to inconsistency in the data and missing crucial information that is not contained in the input data. Consequently, adding additional data such as yield class or age could improve the results.

By performing an ablation study, we examined/determined the importance of the individual input features. The results show that particularly orthophotos contain the most critical information for the model that assesses thinnings. Adding the CHM further improved the performance predicting thinnings. Hence we conclude that the CHM contains unique information that the model could not derive from the orthophotos, whereas the input features DTM and Slope seem not to contain any additional helpful information.

As stated earlier, the proposed model is specifically trained to detect the need of selective crown thinnings in spruce dominated forest stands in the study area. Consequently, further research should examine the feasibility of employing DCNNs for other thinning types, additional tree species and other areas. Particularly the training of a comparable model for deciduous forests seems challenging. The higher diversity of tree species and the more difficult task of segmenting deciduous tree crowns make it an ambitious problem to solve.

Moreover, the performance of assigning the urgency of thinnings was unsatisfying in this study. Hence we propose to conduct further research on this objective by providing additional valuable data such as age or yield class. Another potential direction of research can be the prediction of the volume of harvested wood for sales planning.

In this study, we showed the ability to detect the necessity of thinnings in spruce forests through remote sensing data. Whereas the resulting model needs further tuning for production, we showed the potential of using remote sensing data to plan thinnings cost-effectively. Especially for small forest owners with limited funds, but also for forest management of companies as a quick help, this approach provides the opportunity to receive critical information about the forest promptly.

Bibliography

Abetz, P. (1970). Biologische produktionsmodelle als entscheidungshilfen im waldbau. *Forstarchiv*, (41):5–6.

Avery, T. E. (1966). *Forester's guide to aerial photo interpretation*. Number 308. US Department of Agriculture, Forest Service.

Axelsson, A., Lindberg, E., Reese, H., and Olsson, H. (2021). Tree species classification using Sentinel-2 imagery and Bayesian inference. *International Journal of Applied Earth Observation and Geoinformation*, 100(April 2020):102318.

Ballanti, L., Blesius, L., Hines, E., and Kruse, B. (2016). Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sensing*, 8(6):1–18.

basemap (2021). Orthofoto tilecache of austria, published by geoland.at.

Bohlin, J., Bohlin, I., Jonzén, J., and Nilsson, M. (2017). Mapping forest attributes using data from stereophotogrammetry of aerial images and field data from the national forest inventory. *Silva Fennica*, 51(2):1–18.

Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., and Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment*, 112(10):3876–3890.

Cameron, A. D. (2002). Importance of early selective thinning in the development of long-term stand stability and improved log quality: A review. *Forestry*, 75(1):25–35.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2018a). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018b). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11211 LNCS:833–851.

Chen, Z., Wang, C., Li, J., Fan, W., Du, J., and Zhong, B. (2021). Adaboost-like End-to-End multiple lightweight U-nets for road extraction from optical remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 100:102341.

Cooper, S., Okujeni, A., Pflugmacher, D., van der Linden, S., and Hostert, P. (2021). Combining simulated hyperspectral EnMAP and Landsat time series for forest aboveground biomass mapping. *International Journal of Applied Earth Observation and Geoinformation*, 98(February):102307.

Daume, S. and Robertson, D. (2000). A heuristic approach to modelling thinnings. *Silva Fennica*, 34(3):237–249.

Diakogiannis, F. I., Waldner, F., Caccetta, P., and Wu, C. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162(March 2019):94–114.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., and Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186:64–87.

Fricker, G. A., Ventura, J. D., Wolf, J. A., North, M. P., Davis, F. W., and Franklin, J. (2019). A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery. *Remote Sensing*, 11(19).

Ganz, S., Käber, Y., and Adler, P. (2019). Measuring tree height with remote sensing-a comparison of photogrammetric and LiDAR data with different field measurements. *Forests*, 10(8).

Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., and Plaza, A. (2017). Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):37–78.

Gruber, M., Ponticelli, M., Ladstädter, R., and Wiechert, A. (2012). Ultracam Eagle, Details and Insight. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B1(September):15–19.

Haara, A. and Korhonen, K. T. (2004). Toimenpide-ehdotusten simulointi laskennallisesti ajantasaistetusta kuvioaineistosta. *Metsätieteen aikakauskirja*, 2004(2):157–173.

Halme, E., Pellikka, P., and Mõttus, M. (2019). Utility of hyperspectral compared to multispectral remote sensing data in estimating forest biomass and structure variables in Finnish boreal forest. *International Journal of Applied Earth Observation and Geoinformation*, 83(April):101942.

He, K. and Sun, J. (2015). Convolutional neural networks at constrained time cost. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June-2015(3):5353–5360.

He, K., Zhang, X., Ren, S., and Sun, J. (2015a). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision*, 2015 Inter:1026–1034.

He, K., Zhang, X., Ren, S., and Sun, J. (2015b). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1904–1916.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:770–778.

Hein, S., Herbstritt, S., and Kohnle, U. (2008). Auswirkung der Z-Baum-Auslesedurchforstung auf Wachstum, Sortenertrag und Wertleistung im Europäischen Fichten-Stammzahlversuch (Picea Abies [L.] Karst.) in Südwestdeutschland. *Allgemeine Forst- und Jagdzeitung*, 179(10-11):192–201.

Holgén, P., Söderberg, U., and Hånell, B. (2003). Diameter increment in Picea abies shelterwood stands in northern Sweden. *Scandinavian Journal of Forest Research*, 18(2):163–167.

Holopainen, M., Vastaranta, M., and Hyyppä, J. (2014). Outlook for the next generation's precision forestry in Finland. *Forests*, 5(7):1682–1694.

Hordiiuk, D., Oliinyk, I., Hnatushenko, V., and Maksymov, K. (2019). Semantic segmentation for ships detection from satellite imagery. In *2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO)*, pages 454–457. IEEE.

Horn, B. K. (1981). Hill Shading and the Reflectance Map. *Proceedings of the IEEE*, 69(1):14–47.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua:2261–2269.

Hynynen, J., Ahtikoski, A., Siitonen, J., Sievänen, R., and Liski, J. (2005). Applying the MOTTI simulator to analyse the effects of alternative management schedules on timber and non-timber production. *Forest Ecology and Management*, 207(1-2 SPEC. ISS.):5–18.

Hyvönen, P. (2002). Kuvioittaisten puustotunnusten ja toimen- pide-ehdotusten estimointi k-lähimmän naapurin menetel- mällä Landsat TM -satelliittikuvan, vanhan inventointitiedon ja kuviotason tukiaineiston avulla [Estimation of stand characteristics and forest management op. *Metsätieteen aikakauskirja*, 2(2002):363–379.

Hyvönen, P., Heinonen, J., and Haara, A. (2010). Detection of forest management operations using Bi-temporal aerial photographs. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 38(1999):309–313.

Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., and Atzberger, C. (2019). Optimal input features for tree species classification in Central Europe based on multi-temporal Sentinel-2 data. *Remote Sensing*, 11(22).

Jadon, S. (2020). A survey of loss functions for semantic segmentation. *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2020*.

Jegou, S., Drozdzal, M., Vazquez, D., Romero, A., and Bengio, Y. (2017). The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2017-July:1175–1183.

Juodvalkis, A., Kairiukstis, L., and Vasiliauskas, R. (2005). Effects of thinning on growth of six tree species in north-temperate forests of Lithuania. *European Journal of Forest Research*, 124(3):187–192.

Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K. T., Maltamo, M., Nilsson, M., Nord-Larsen, T., Næsset, E., and Olsson, H. (2018). Remote sensing and forest inventories in Nordic countries–roadmap for the future. *Scandinavian Journal of Forest Research*, 33(4):397–412.

Kingma, D. P. and Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–13.

Liu, M., Fu, B., Xie, S., He, H., Lan, F., Li, Y., Lou, P., and Fan, D. (2021). Comparison of multi-source satellite images for classifying marsh vegetation using DeepLabV3 Plus deep learning algorithm. *Ecological Indicators*, 125:107562.

Liu, Y., Gong, W., Xing, Y., Hu, X., and Gong, J. (2019). Estimation of the forest stand mean height and aboveground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 151(March):277–289.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 07-12-June, pages 3431–3440. IEEE Computer Society.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152(April):166–177.

Macdonald, E., Gardiner, B., and Mason, W. (2010). The effects of transformation of even-aged stands to continuous cover forestry on conifer log quality and wood properties in the UK. *Forestry*, 83(1):1–16.

MacKenzie, R. F. (1976). Silviculture and management in relation to risk of windthrow in Northern Ireland. *Irish Forestry*, 33:29–38.

Magnussen, S., Nord-Larsen, T., and Riis-Nielsen, T. (2018). Lidar supported estimators of wood volume and aboveground biomass from the Danish national forest inventory (2012–2016). *Remote Sensing of Environment*, 211(January):146–153.

Mitchell, S. J. (2000). Stem growth responses in Douglas-fir and sitka spruce following thinning: Implications for assessing wind-firmness. *Forest Ecology and Management*, 135(1-3):105–114.

Olesk, A., Voormansik, K., Põhjala, M., and Noorma, M. (2015). Forest change detection from Sentinel-1 and ALOS-2 satellite images. *Proceedings of the 2015 IEEE 5th Asia-Pacific Conference on Synthetic Aperture Radar, APSAR 2015*, pages 522–527.

Persson, P. (1975). Windthrow in forests-its causes and the effect of forestry measures [sweden, scots pine, norway spruce]. *Rapporter och Uppsatser-Skogshoegskolan, Institutionen foer Skogsproduktion (Sweden)*.

Pollanschutz, J. (1980). Erfahrung aus der schneebruchkatastrophe 1979. *Allegem. Forstztg.*, 91(5):123–125.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9351, pages 234–241. Springer Verlag.

Schütz, J. P., Götz, M., Schmid, W., and Mandallaz, D. (2006). Vulnerability of spruce (Picea abies) and beech (Fagus sylvatica) forest stands to storms and consequences for silviculture. *European Journal of Forest Research*, 125(3):291–302.

Simard, M., Pinto, N., Fisher, J. B., and Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences*, 116(4):1–12.

Slodicak, M., Novak, J., and Skovsgaard, J. P. (2005). Wood production, litter fall and humus accumulation in a Czech thinning experiment in Norway spruce (Picea abies (L.) Karst.). *Forest Ecology and Management*, 209(1-2):157–166.

Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 4:2951–2959.

Spellmann, H. and Schmidt, M. (2003). Massen-, sorten-und wertertrag der fichte in abhangigkeit von der bestandesbehandlung. *Forst und Holz*, 58(13/14):412–419.

Stirling, G., Gardiner, B., Connolly, T., Mochan, S., and Macdonald, E. (2000). A survey of Sitkka spruce Stem Straightness in South Scotland. (September).

Stoian, A., Poulain, V., Inglada, J., Poughon, V., and Derksen, D. (2019). Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. *Remote Sensing*, 11(17):1–26.

Valinger, E. and Fridman, J. (1999). Models to assess the risk of snow and wind damage in pine, spruce, and birch forests in Sweden. *Environmental Management*, 24(2):209–217.

Valinger, E., Lövenius, M. O., Johansson, U., Fridman, J., Claeson, S., and Gustafsson, Å. (2006). Analys av riskfaktorer efter stormen Gudrun.

Vastaranta, M., Holopainen, M., Yu, X., Hyyppä, J., Hyyppä, H., and Viitala, R. (2011). Predicting stand-thinning maturity from airborne laser scanning data. *Scandinavian Journal of Forest Research*, 26(2):187–196.

Volpi, M. and Tuia, D. (2018). Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144(July):48–60.

Yan, S., Jing, L., and Wang, H. (2021). A new individual tree species recognition method based on a convolutional neural network and high-spatial resolution remote sensing imagery. *Remote Sensing*, 13(3):1–21.

Yu, X., Hyyppä, J., Kaartinen, H., and Maltamo, M. (2004). Automatic detection of harvested trees and determination of forest growth using airborne laser scanning. *Remote Sensing of Environment*, 90(4):451–462.

Yue, K., Yang, L., Li, R., Hu, W., Zhang, F., and Li, W. (2019). TreeUNet: Adaptive Tree convolutional neural networks for subdecimeter aerial image segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 156(April 2018):1–13.

Appendix A

Final map

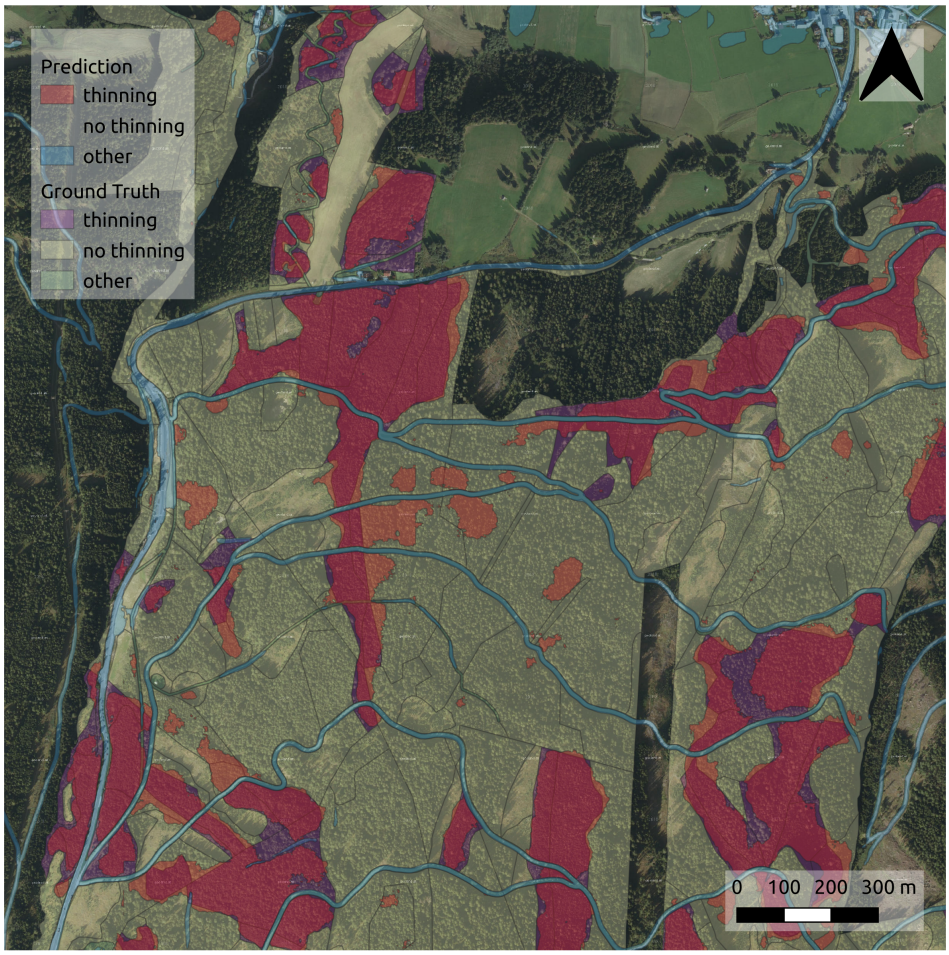


Figure A.1: Prediction of the necessity of thinning of the final model *Base* in part of the study area of Lungau, Austria. The background image is a true colour orthophoto from airborne photography.

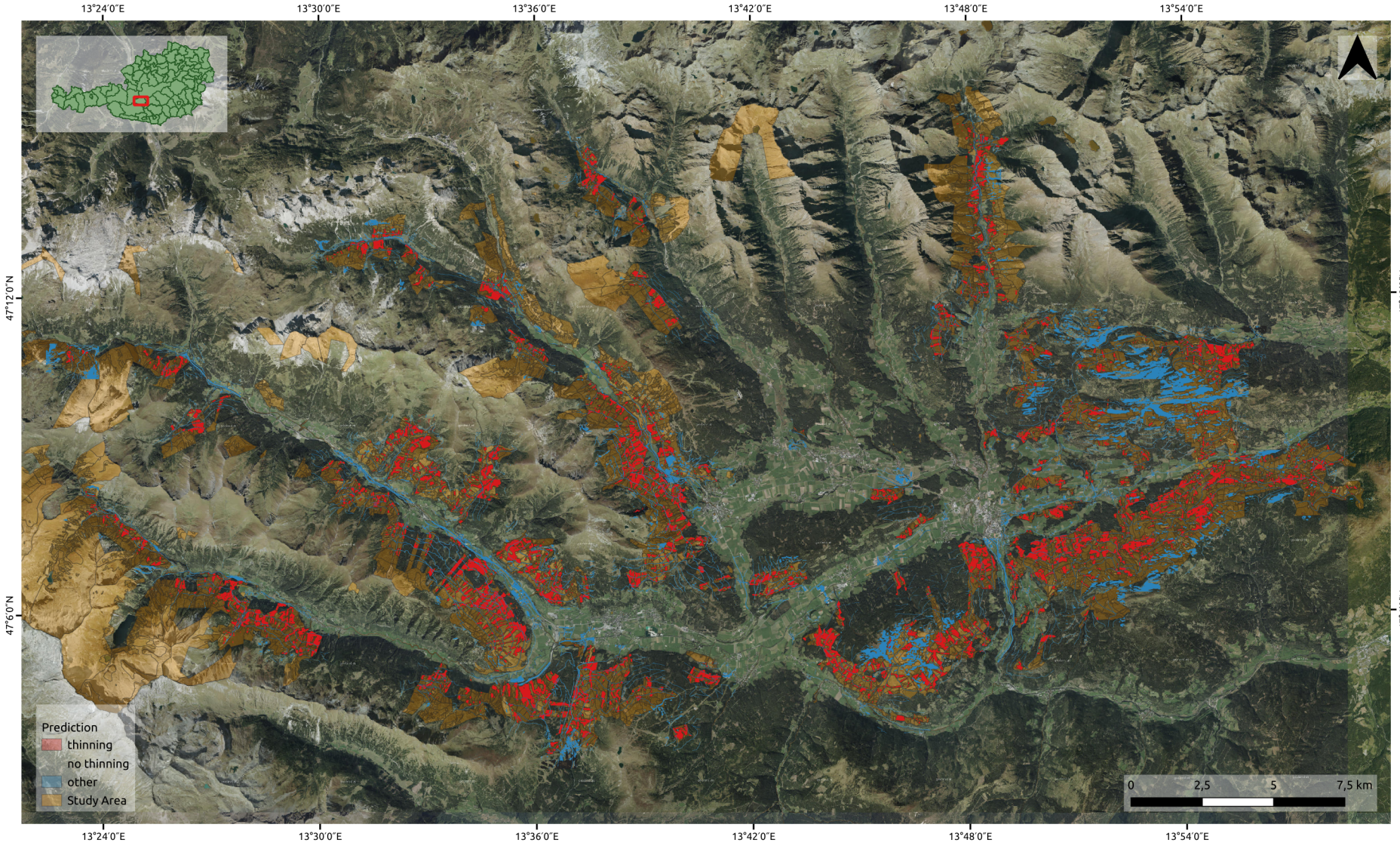


Figure A.2: Prediction of the necessity of thinning of the final model *Base* in the study area of Lungau, Austria. The background image is a true colour orthophoto from airborne photography.

Appendix B

Final DCNN architecture

Figure B.1 shows the final DCNN-architecture tuned on the data set (ground truth *Base*). We applied this architecture to train all models to resolve the research objectives ”thinning necessity” (Subsection 5.2.1), ”thinning urgency” (Subsection 5.2.2) and the ablation study (section 5.3).

The diagram (Figure B.1) illustrates the overall structure of the network. However, batch norm layers and relu activation functions are omitted due to space constraints. A batch norm layer follows every convolution as well as atrous convolution. Whereas the relu activation functions follow every Resnet-101 convolution block (fine dotted violet lines), otherwise every convolution and atrous convolution. Furthermore, the numbers on the right side of the ResNet-101 blocks express how many times the block is repeated. The information flow into the Connection happens after the first Resnet-101 block is repeated three times. The PyTorch implementation is provided at `https://github.com/satlawa/edin_thinning_necessity`.

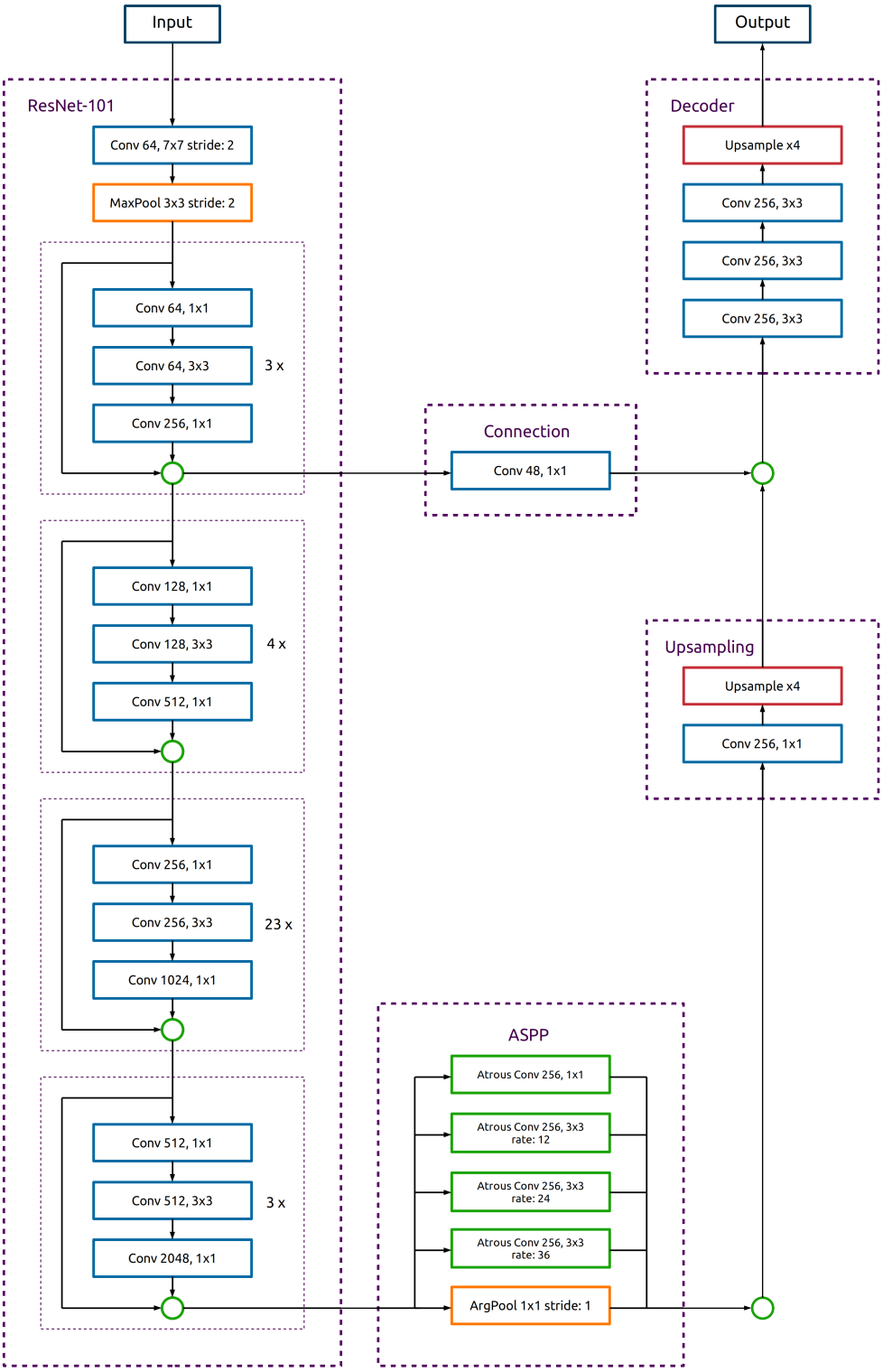


Figure B.1: Diagram of DeepLabv3+ with ResNet-101 as backbone (Chen et al. (2018b)). This network architecture is the final deep convolutional neural network used to train all models to detect the necessity and the urgency of thinnings. Green circles: concatenations.

Appendix C

DCNN architectures

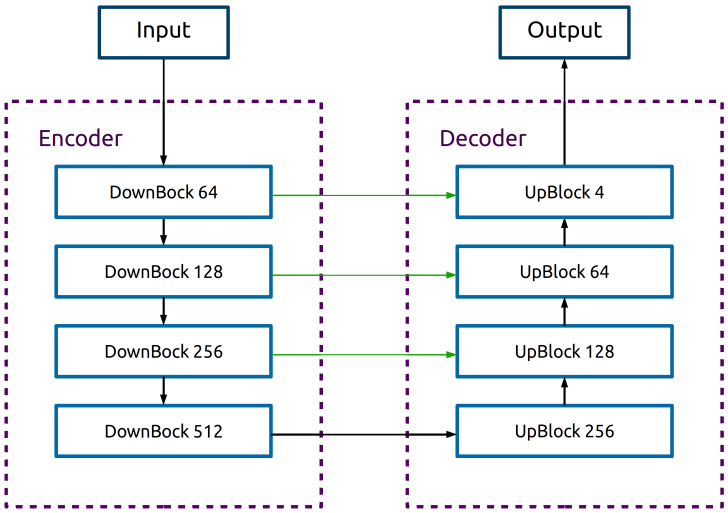


Figure C.1: Diagram of modified UNet used in this study. The architecture is composed of Downsampling Blocks (DownBlock) and Upsampling Blocks (UpBlock). Table C.1 shows a detailed composition of the blocks. The number inside the blocks illustrates the number of feature maps used in the conulution layers.

Table C.1: Composition of the Downsampling and Upsampling Block of the UNet.

Downsampling Block (DownBlock)	**Upsampling Block (UpBlock)**
Convolution 3x3, stride 1 Batch Normalisation ReLU	Convolution 3x3, stride 1 Batch Normalisation ReLU
Convolution 3x3, stride 1 Batch Normalisation ReLU	Convolution 3x3, stride 1 Batch Normalisation ReLU
MaxPool 3x3, stride 2	Transposed Convolution 3x3, stride 2

Table C.2: Composition of the main building blocks of FC-DensNet.

Layer	**Transition Down (TD)**	**Transition Up (TU)**
Batch Normalization ReLU Convolution 3 x 3 Dropout p = 0.2	Batch Normalization ReLU Convolution 1 × 1 Dropout p = 0.2 Max Pooling 2 × 2	Transposed 3 × 3, Convolution, stride = 2

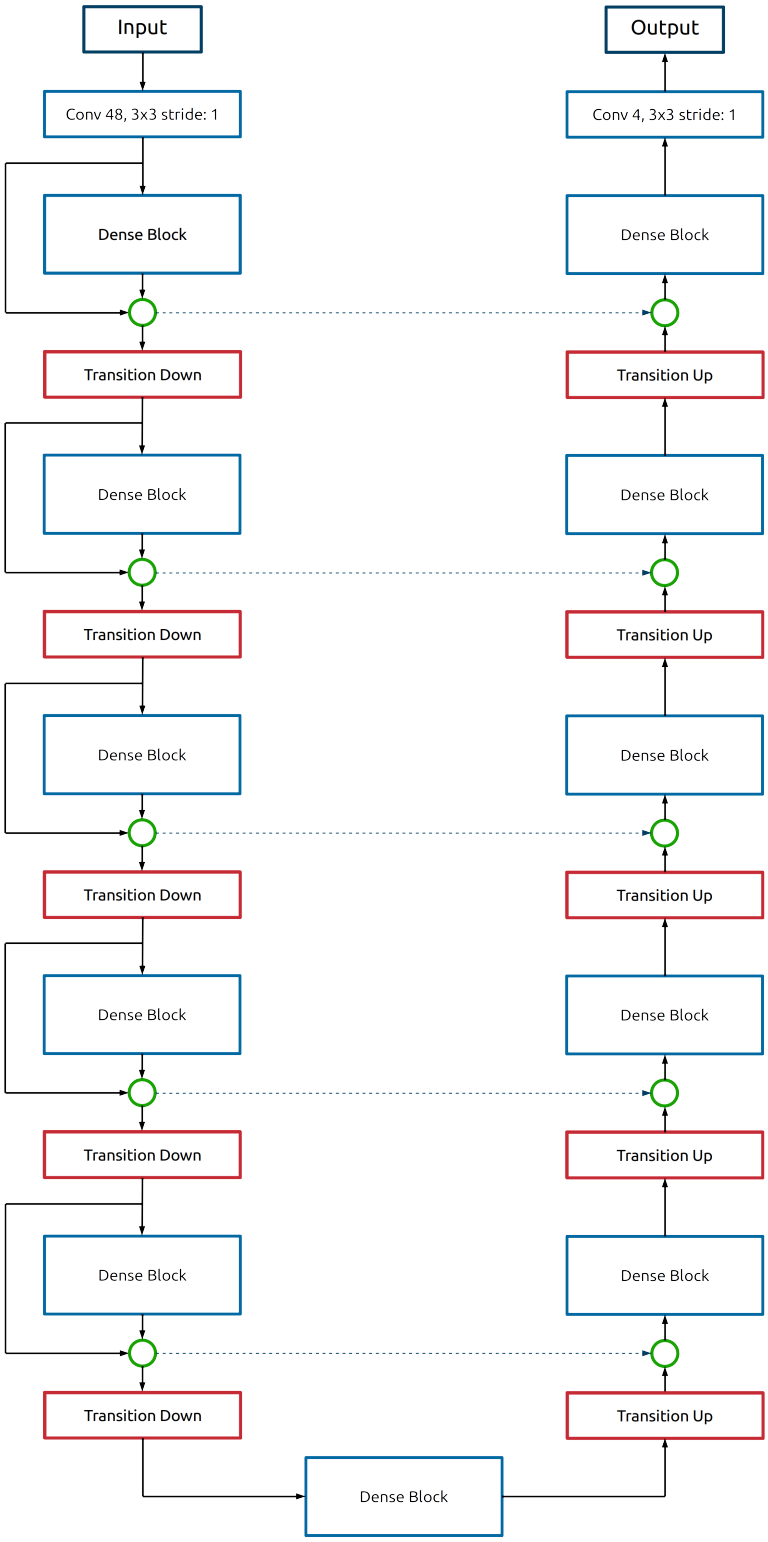


Figure C.2: Diagram of FC-DenseNet56 (Jegou et al. (2017)) used in this study. The architecture is composed of Dense, Transition Up and Transition Down Blocks. Table C.2 shows a detailed composition of the blocks. The network architecture was not modified (except for the input and output dimensions) and is exactly the same as in Jegou et al. (2017) (FC-DenseNet56).Green circles: concatenations.

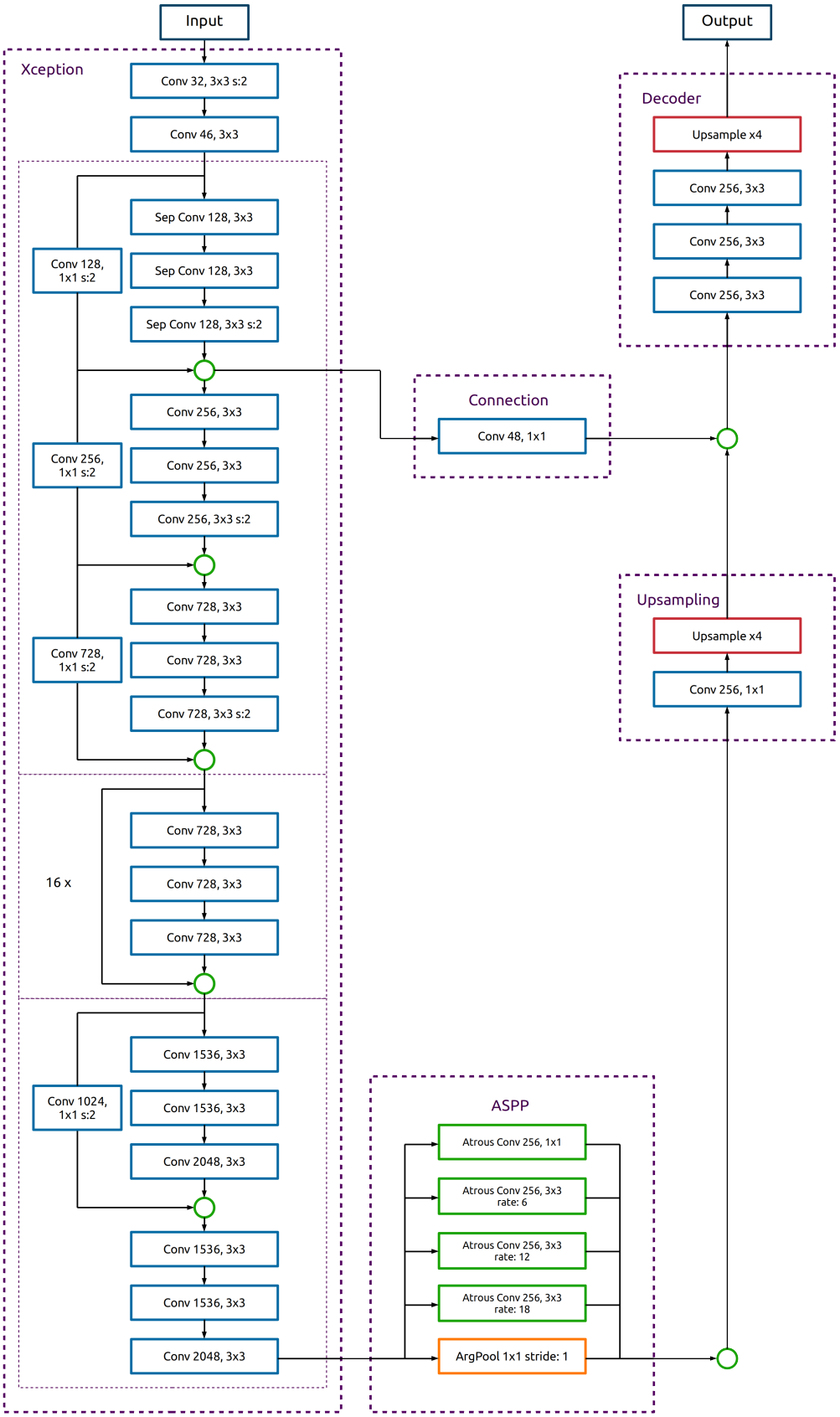


Figure C.3: Diagram of DeepLabv3+ with Xception as backbone as originally proposed in Chen et al. (2018b). The network architecture was not modified (except for the input and output dimensions). Green circles: concatenations.

Appendix D

Additional results

Here we provide the confusion matrices and scores of selected models that were omitted from the main part of this study.

D.1 Confusion matrices

Table D.1: Confusion matrix on test set (ground truth *masked UR12*). The numbers represent classified pixels in 10^6.

		prediction			
	Class	**thinning ur1**	**thinning ur2**	**no thinning**	$\sum$
reference	**thinning ur1**	31	11	6	49
	thinning ur2	25	32	8	65
	no thinning	17	20	183	220
	$\sum$	74	63	196	333

D.2 Scores

Here we present the individual scores of the 5-fold cross validation in Subsection 5.2.1.

Table D.2: Confusion matrix on test set (ground truth *UR1*). The numbers represent classified pixels in 10^6.

		prediction			
	Class	**thinning ur1**	**merged**	**other**	$\sum$
reference	**thinning ur1**	27	22	0	48
	merged	30	399	6	435
	other	0	8	24	32
	$\sum$	57	429	30	516

Table D.3: Confusion matrix on test set (ground truth *masked UR1*). The numbers represent classified pixels in 10^6. ref: reference.

		prediction		
	Class	**thinning ur1**	**merged**	$\sum$
ref	**thinning ur1**	30	18	48
	merged	37	248	285
	$\sum$	67	266	333

Table D.4: Class scores and mean class scores for fold-1 on the test set (ground truth *Base*).The model's objective is to detect the necessity of thinning.

Score	**thinning**	**no thinning**	**other**	**mean**
precision	75.97%	92.08%	77.35%	81.8%
recall	82.45%	89.76%	75.14%	82.45%
IoU	65.39%	83.32%	61.59%	70.1%
F1	79.08%	90.9%	76.23%	82.07%

Table D.5: Class scores and mean class scores for fold-2 on the test set (ground truth *Base*).The model's objective is to detect the necessity of thinning.

Score	thinning	no thinning	other	mean
precision	75.26%	92.01%	79%	82.09%
recall	80.76%	90.36%	74.29%	81.8%
IoU	63.82%	83.78%	62.04%	69.88%
F1	77.91%	91.17%	76.58%	81.89%

Table D.6: Class scores and mean class scores for fold-3 on the test set (ground truth *Base*).The model's objective is to detect the necessity of thinning.

Score	thinning	no thinning	other	mean
precision	78.11%	91.3%	81.27%	83.56%
recall	79.83%	91.51%	73.35%	81.56%
IoU	65.24%	84.17%	62.75%	70.72%
F1	78.96%	91.4%	77.11%	82.49%

Table D.7: Class scores and mean class scores for fold-4 on the test set (ground truth *Base*).The model's objective is to detect the necessity of thinning.

Score	thinning	no thinning	other	mean
precision	77.3%	91.05%	81.42%	83.25%
recall	79.65%	91.15%	72.45%	81.08%
IoU	64.54%	83.65%	62.17%	70.12%
F1	78.45%	91.1%	76.68%	82.08%

Table D.8: Class scores and mean class scores for fold-5 on the test set (ground truth *Base*).The model's objective is to detect the necessity of thinning.

Score	thinning	no thinning	other	mean
precision	78.76%	91.44%	79.39%	83.2%
recall	78.89%	91.82%	75.5%	82.07%
IoU	65.05%	84.55%	63.13%	70.91%
F1	78.82%	91.63%	77.4%	82.62%