Mapping Land cover for Finland

18.04.2023

Jonas Tranberg, jonas@scalgo.com

Helena Åström, helena@scalgo.com

Table of contents

1 Introduction 3

1.1 Project description 3

1.2 Background 3

1.3 Project phases 5

1.3.1 Improve the UNET-model 5

1.3.2 Include auxiliary data 7

1.4 Final product 7

2 Method 8

2.1 Data Sources 8

2.2 Description of UNET model 10

2.3 Post processing 10

2.3.1 Auxiliary data 10

2.3.2 Raster map 11

3 Validation 12

3.1 Validation area examples 12

1 Introduction

This documentation describes the production of a land cover map for Finland, ordered by the Finnish Environment Institute (SYKE). The report includes a description of methodology and a validation of the results.

1.1 Project description

The goal of the project was to produce a land cover map for the entire Finland using a Machine learning model that identifies land cover from aerial images. Results from the Machine learning were combined with auxiliary data to further detail some of the land cover classes.

A UNET-based Convolutional Neural network model used in this project, is developed by SCALGO in several previous projects and in close collaboration with SYKE. Main input data used in this project is Aerial images provided by Maanmittauslaitos and auxiliary data of roads, buildings and fields. The final land cover classes are:

Main class	**Sub classes**
Impervious areas	Buildings Paved roads Other impervious
Green surfaces	Dense vegetation Shallow vegetation Fields
Bare land	Unpaved roads Other bare land
Bare rock	
Water	Fresh water Sea

1.2 Background

The UNET-based model was first developed by SCALGO and Aarhus University in 2018-2019. The initial model was tuned for separating impervious areas from previous surfaces, i.e., separating sand and gravel from asphalt. This classification is, with traditional methods, very difficult.

To train the initial model, data from Danish Water utilities and consultants were used. In 2019, SCALGO produced the first nation-wide land cover map for Denmark, see Figure 1 for an example.

Figure 1. Land cover map produced for Denmark. The map separates previous surfaces (green shades) from impervious surfaces (orange and red).

The UNET model is improved in several projects to better describe Finnish conditions:

- In the LaserVesi-project, financed by Maa- ja Metsätalousministeriö, the UNET model was tested and further developed with the goal to produce an imperviousness map in the Helsinki region. The project was conducted in collaboration with HSY, City of Helsinki and SYKE.

- As a continuation of the LaserVesi project, a smaller test was made by HSY and SCALGO to evaluate if the model can identify bare land (sand and gravel). Hence, the model was further trained to conduct this classification.

- In autumn 2022, the model was used to produce “Helsingin seudun maanpeiteaineisto” in collaboration with HSY. Here, many more classes, such as bare rock and vegetation classes, were introduced to the model successfully.

All above mentioned maps are freely available in the HSY map service.

1.3 Project phases

This project consisted of two main phases:

1. Improve the UNET-model performance at national scale
2. Include auxiliary data to produce a map with all classes mentioned in Table 1

1.3.1 Improve the UNET-model

The model, at the start of this project, was especially tuned for classifying land cover in Southern Finland because of the LaserVesi project. However, to produce land cover classification for the entire Finland, the model must be further trained to understand the variation in quality and color of aerial images. Depending on when and by whom the images are collected, there can be considerable dissimilarities, making it difficult for the model to identify land cover. For example, taking images in spring or summer makes land cover look very different, as shown in Figure 2.

Figure 2. Left: aerial image taken in the spring, Right: aerial image taken in the summer.

The landscape in Finland also differs depending on where in the country you are, making it important that the UNET-model can analyze various types of landscapes. Figure 3 shows examples from Northern Finland with large wilderness and limited tree cover, Eastern Finland with dense forest cover, and Western Finland with agriculture.

In this project, the model was further trained with data from all over Finland.

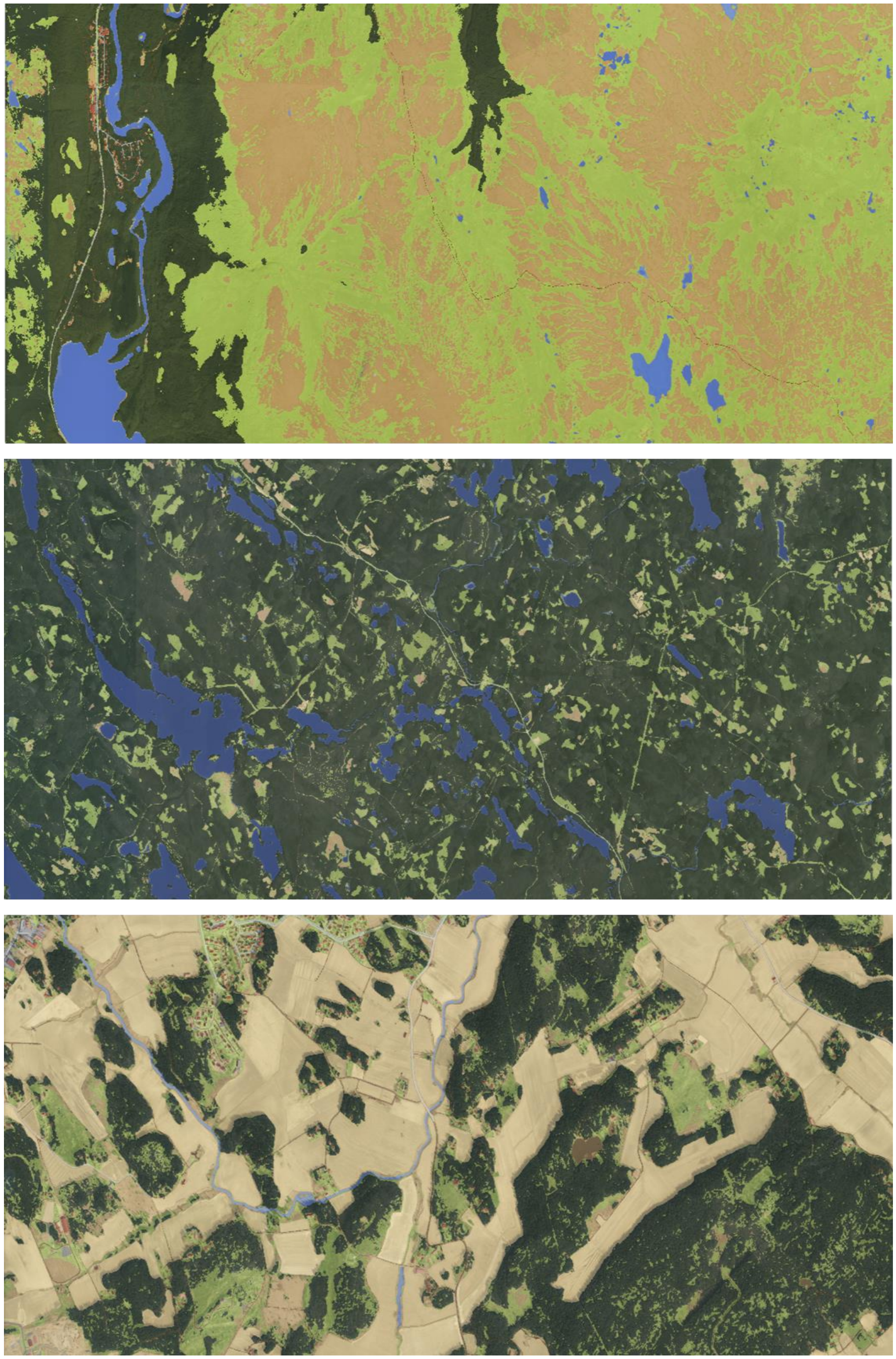

Figure 3. Top: Northern Finland, Middle: Eastern Finland, Bottom: Western Finland

1.3.2 Include auxiliary data

In the second phase of the project, standard GIS tools were used to integrate auxiliary data into the land cover map. Following auxiliary data was included:

- Fields
- Buildings
- Roads (both unpaved and paved)

Fields and Buildings were included as existing polygons. Roads were included using road centerlines from Digiroad.

1.4 Final product

In the final product, land cover classes are defined with the following codes and colors:

Land cover class	**Code**	**RGB colors**
Paved road	111	(130, 130, 130),
Unpaved road	112	(255, 217, 128),
Building	120	(178, 223, 67),
Other impervious	130	(50, 70, 40),
Field	211	(189, 189, 189),
Shallow vegetation	212	(130, 69, 19),
Dense vegetation	220	(216, 71, 41),
Bare rock	310	(208, 202, 208),
Bare land	410	(209, 161, 82),
Fresh water	510	(100, 149, 237),
Sea	520	(100, 149, 237),

The data is produced in the ETRS-TM35FIN coordinate system, in GeoTIFF format . The data is available in two resolutions, 0,5 m and 2 m.

2 Method

This chapter details the method for producing a nation-wide land cover map.

2.1 Data Sources

Below is a table of the auxiliary data sources used to produce the landcover map. Further processing of these sources is described in the *post process* section.

Name	Description	Date of download
RGB	RGB images from http://tiedostopalvelu.maanmittauslaitos.fi , used as the main input for the machine learning model. Data was flown between 2017 and 2022.	2022-10-31
Roads	Paved and unpaved [1] centerlines of roads from https://vayla.fi/en/transport-network/data/digiroad/data.	2023-02-02
Buildings	From maanmittauslaitos. https://tiedostopalvelu.maanmittauslaitos.fi/tp/feed/mtp/maastotietokanta/kaikki. All buildings expect those with LUOKKA = 44300	2023-01-29

[1] We create **paved** and **unpaved** road polygons from DigiRoad center lines. We first join the tables DR_LINKKI_K, DR_PAALLYSTETTY_TIE_K and DR_LEVEYS_K. We determine whether it's paved or unpaved using the *paal_arvo* property (from the table DR_PAALLYSTETTY_TIE_K), if it does not exist then we use the road type property (*toiminn_lk)*, and determine that the road paved if this property is less than 5. We filter roads below the surface (*silte_alik* < *0*). Then we buffer each line string feature, with a buffer width is set by the following:

If the road width property (*leve_arvo*, from the table DR_LEVEYS_K) exist, then we use that value, if it doesn't exist we use the following heuristic to get the road width:

- If roadtype=1-> 10m
- If roadtype=2-> 8m
- If roadtype=3-> 7m
- If roadtype=4-> 6m
- If roadtype=5-> 5m
- If roadtype=6 or 7-> 4m

Water	From maanmittauslaitos. https://tiedostopalvelu.maanmittauslaitos.fi/tp/feed/mtp/maastotietokanta/kaikki. LUOKKA >= 36000 and LUOKKA != 36211 and LUOKKA < 36390	2023-01-29
Farmland	From maanmittauslaitos. https://tiedostopalvelu.maanmittauslaitos.fi/tp/feed/mtp/maastotietokanta/kaikki. LUOKKA = 32611	2023-01-29
Sea	From maanmittauslaitos. https://tiedostopalvelu.maanmittauslaitos.fi/tp/feed/mtp/maastotietokanta/kaikki. LUOKKA = 36211	2023-01-29
Railroad bridges	From maanmittauslaitos. https://tiedostopalvelu.maanmittauslaitos.fi/tp/feed/mtp/maastotietokanta/kaikki. ` LUOKKA >=14110 and LUOKKA <=14140	2023-01-29

2.2 Description of UNET model

A UNET-like Convolutional Neural network was trained using training data generated from internal processes, previous projects and in collaboration with SYKE.

UNET has become a standard when using a training-based approach to segment raster data. The advantage that UNET, or most other CNN based architectures, has against other learning-based methods such as *Random Forests* or *Gradient Boosted Trees*, is its ability to integrate the context of the pixel neighborhood using convolution operations at multiple scales through maxpool operations into the prediction of single pixel output in an efficient manner.

Input to the model consists of RGB Images supplied by Maanmittauslaitos[2] and Road centerlines from Digiroad[3].

The output from the Machine learning model consists of 5 district outputs per pixel expressing the probability of each class. These classes are:

- Impervious
- Bare Land
- Shallow vegetation
- Dense vegetation
- Water
- Bare rock

The output from the UNET-model is not used directly, but it informs the final classes in various ways as described below.

2.3 Post processing

In addition to the Machine learning intermediate outputs, we create a set of auxiliary outputs and, thereafter, construct a raster map. These steps are described below.

2.3.1 Auxiliary data

Creating auxiliary data includes the following steps:

[2] https://tiedostopalvelu.maanmittauslaitos.fi/tp/feed/mtp, 10-2022

[3] https://vayla.fi/en/transport-network/data/digiroad/data, 10-2022

- We create **paved** and **unpaved** road polygons from DigiRoad center lines. We determine whether it's paved or unpaved using the *paal_arvo* property, if it does not exist then we use the road type property (*toiminn_lk)* and determine that the road paved if this property is less than 5. We filter roads below the surface (*silte_alik* < *0*). Then we buffer each line string feature, with a buffer width is set by the following:

  If the road width property (*leve_arvo*) exist, then we use that value, if it doesn't exist, we use the following heuristic to get the road width:

  - If $roadtype = 1 -> \ 10m$
  - If $roadtype = 2 -> \ 8m$
  - If $roadtype = 3 -> \ 7m$
  - If $roadtype = 4 -> \ 6m$
  - If $roadtype = 5 -> \ 5m$
  - If $roadtype = 6 \ or \ 7 -> \ 4m$

- We create additional **other-impervious** areas from the set of railway centerlines attributed as bridges. That is, we buffer these centerlines with 3.75 meters (corresponding to a track width of 7.5 meters).

- We gather **field**, **building**, **water** polygons from MML

2.3.2 Raster map

Based on the above-mentioned sources and the Machine Learning probability rasters, we construct the categorical raster for every pixel, using the following rules applied in the specified order:

1. We set the output to the Machine Learning class with the highest probability.

2. If a pixel is inside a water polygon, we set it to the water class.

3. If a pixel is inside a field polygon and is not currently classified as *dense vegetation,* we classify it as field.

4. If a pixel is inside a paved road polygon and the pixel is not currently categorized as *shallow vegetation*, we set it to the paved road class.

5. If a pixel is inside an unpaved road polygon, we classify it as such.

6. If a pixel is inside a building polygon, we classify it as such.

3 Validation

To validate the results, we evaluate performance on 33 separate polygons drawn and labeled by SCALGO and SYKE compromising 38.125.030 pixels in total. We evaluate the machine learning classes, instead of the final classes, as great performance on the auxiliary data is per definition trivial.

The confusion matrix[4] for the classes are displayed below:

	IV	BL	SV	DV	WA	BR
IV	**0.976915**	0.000000	0.023085	0.000000	0.000000	0.000000
BL	0.013447	**0.954031**	0.032461	0.000061	0.000000	0.000000
SV	0.000012	0.154852	**0.843961**	0.001175	0.000000	0.000000
DV	0.000223	0.000968	0.065129	**0.933681**	0.000000	0.000000
WA	0.001470	0.030483	0.016751	0.000423	**0.950872**	0.000000
BR	0.001294	0.025644	0.116562	0.000000	0.000000	**0.856501**

In the matrix, we use the following abbreviations: Impervious: (IV), bare land (BL), bare rock (BR), shallow vegetation (SV), dense vegetation (DV), water (WA).

According to the validation, bare land and impervious areas perform well (over 0.95), whereas shallow vegetation and bare rock has a lower performance (approximately 0.85). Hence, when using the data, we suggest that the accuracy of the data is carefully evaluated by the user, and if needed, manually corrected.

3.1 Validation area examples

Examples of polygons from the validation set. True labels for each polygon are, from left to right: *bare-land, impervious, dense-vegetation, shallow-vegetation, bare-land,* and *bare-rock*.

[4] https://en.wikipedia.org/wiki/Confusion_matrix