SCALGO

Documentation for the LaserVesi project

Mapping imperviousness for Greater Helsinki Region

21.02.2022

Helena Åström, helena@scalgo.com

Jonas Tranberg, jonas@scalgo.com

SCALGO

Table of Contents

Introduction

Artificial impervious surfaces increase in cities, and this contributes to challenges with surface water runoff, including capacity issues in drainage systems, increased overflow to recipients, and flooding. Land cover data, describing pervious areas (e.g. bare land, vegetation) and impervious areas (e.g. paved roads, buildings), is important input for surface water and drainage modelling, and decisions taken therefrom.

In 2021, The Ministry of Agriculture and Forestry of Finland financed an innovation project, called LaserVesi, partnering the Finnish Environment institute (SYKE), the Utility company Helsinki Region Environmental Services Authority (HSY), City of Helsinki, and SCALGO that specializes in large scale analysis of geographic data.

The project tests the applicability of new data from the National Laser Scanning and Aerial Imaging Program to support water management. One of the goals in the project, was to test a deep learning model for more mapping imperviousness in Greater Helsinki Region.

Results in this pilot project were produced in the period February 2021-September 2021.

Method

Short description of model

A UNET based “Convolutional Neural Network” model, developed by SCALGO in close cooperation with Aarhus University, was used for mapping imperviousness. Output is presented as probability of imperviousness per pixel.

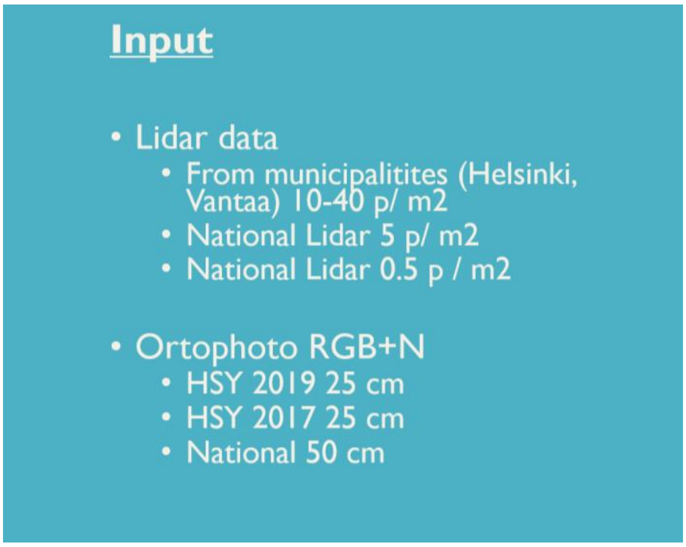
UNET has become a standard when using a training-based approach to segment raster data. The power of UNET, or any other CNN based architecture, over other learning-based methods, is the way that it integrates the context of a single pixel prediction over multiple scales, such that information in the neighbourhood of the predicted pixel, is used in the output prediction in an efficient manner.

Input data

Two different input sources were used:

1. Orthophotos in 4 bands, namely the RGB channels and the Near-Infrared channel
2. A raster-based Lidar point cloud, describing the changes in terrain. That is, we first create 2 rasters from the raw point cloud: A raster based on all the points and a raster based on only ground points, which is then smoothed. Then we subtract the second raster from the first to create a raster describing the high frequency changes in the terrain, while ignoring the general (low frequency) elevation.

In the LaserVesi project, several data sets were tested to evaluate if the choice of input data influences the accuracy. Following data sets were used:



Training data

The model was originally trained with high-quality land cover polygons produced manually by Skanderborg Water Utility. Other contributors of training data include water utilities KLAR, FORS, SAMN, HOFOR, and the engineering company Krüger.

In the LaserVesi project, further training data was acquired to describe the underlying land cover, which are then used to label individual pixels. Among these sources are:

- Polygons derived from the existing Land Cover map of the Greater Helsinki region.
- Annotation made in cooperation with Special Minds in Aarhus, Denmark, covering both areas in Denmark and Finland.

Validation areas

Several validation polygons were used, both during training to gauge the progress of improvement and after training to evaluate the general performance of the model. All these areas were hand-picked by us SCALGO or the LaserVesi project team.

The process

The final model was produced through a series of iterations, each consisting of the following steps.

1. Select a training set
2. Train the model
3. Evaluate the model and modify the training set to cover for errors

This process was repeated until the model was deemed fit for production.

In the figure below is an example of one validation area and how the result developed throughout different iterations:

Output

The model produces a raster where each pixel (of size 0.2m) represents the probability of that pixel being impervious areas. This is also the format, which is delivered to LaserVesi, abait in a single byte resolution resulting in 254 different probability values.

Results

Performance of different data

The project tested the performance of the model with combinations of different data sets. Four different combinations were used:

Data	HLHO	SLHO	SLLSO	SLSO	SLHOS
HSY 2019 orthophotos	x	x			x
HSY 2017 orthophotos	x	x			
MML 50cm orthophotos			x	x	
HSY lidar (Vantaa, Helsinki)	x				
MML lidar high res (5p/m2)		x		x	x
MML lidar low res (0.5p/m2)			x		
Extra SM training data			x	x	x

The performance of these combinations, i.e., how well the model performs on selected validation areas and on average, is presented in the table below:

Validation sets	SLHOS	SLSO	SLLSO	SLHO	HLHO
bare-road-1	0.826146	0.957192	0.946052	0.724115	
helsinki-bare-bare-land-1	0.995883	0.998845	0.999992	0.665924	
helsinki-bare-bare-land-2	0.977454	0.966069	0.995494	0.497285	
helsinki-bare-bare-land-3	0.687629	0.776766	0.7715	0.755569	
helsinki-bare-bare-land-4	0.859525	0.957295	0.956431	0.838789	
helsinki-bare-bare-land-5	0.898011	0.912179	0.930494	0.788675	
helsinki-bare-bare-land-6	0.764951	0.809298	0.782617	0.705408	
helsinki-bare-bare-land-7	0.989168	0.948596	0.985956	0.742166	0.424307
helsinki-other-imp-val-1	0.98353	0.977636	0.952728	0.902256	
helsinki-other-imp-val-2	0.99716	0.999934	0.998648	0.991008	0.847214
helsinki-other-imp-val-3	0.88876	0.903356	0.865141	0.882702	0.932319
paved-road-1	0.981254	0.631425	0.792785	0.986755	
mean	0.904123	0.903216	0.91482	0.790054	
mean_aggr	0.953662	0.952343	0.956703	0.740998	

From the experiments with different data, it was concluded that the most important factors are consistent input features covering large areas, and training data capturing the variance of the terrain. In contrast, resolution of input features is less important.

Performance of final output

The performance of the final output is compared with HSY land cover map. The HSY imperviousness map is created by joining classes, paved road, building and other impervious surfaces. Evaluated on a range of validation areas, created by SCALGO and the project team.

Table 1. Model performance, general sample. Existing land cover is the HSY maanpeiteaineisto, greater-helsinki-3 is the final iteration

	Existing land cover map (HSY)	**greater-helsinki-3**
accuracy	0.780673	0.879181

The comparison shows that greater-helsinki-3 achieves significantly better performance than the existing land cover dataset, even though roads make up a significant portion of the sample and those are human annotated (not learned) in the existing land cover dataset.