

Multisource spatiotemporal land cover classification using the LUISA Basemap and Convolutional Neural Networks

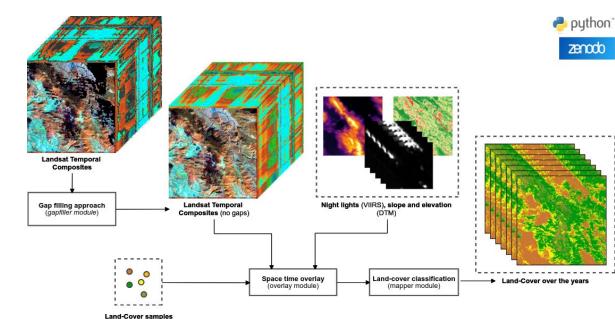
Martijn Witjes martijn.witjes@opengeohub.org

1 Who am I and what do I do?



WAGENINGEN UNIVERSITY & RESEARCH

- PhD student at Wageningen University
- Work at OpenGeoHub
- ML for land use / land cover mapping
- Data wrangling



(LUCAS and CORINE)

What is OpenGeoHub?



Non-profit research foundation

Open Data

Open Source

Education



Datacets - 1 min read

Seamless 30 meter Sentinel-2 L2A Pan-European seasonal cloudless mosaics (2018 - 2020)



Datasets : 1 min read

Presence-Absence Points for Tree Species Distribution Modelling for Europe



Datasets • 1 min read

Potential and realized distribution at 30m for 16 forest tree species in Europe for 2000 -2020



MOOD: MOnitoring Outbreak events for Disease surveillance in a data science context

EU-FUNDED, Training





My research

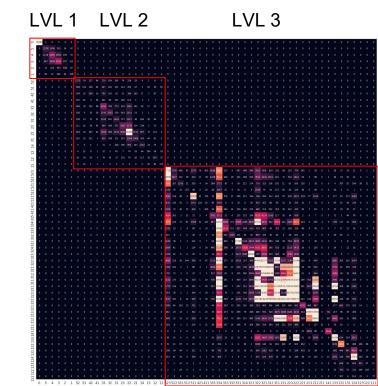


Open data LULC maps that are fit for as many use cases as possible

This leads to:

- 1. Continuity & time series
- 2. Large hierarchical legends
- 3. Optimizable for specific classes

HIERARCHICAL STRUCTURE VISUAL



My Publication



Replicate CORINE LULC annually

2000-2019

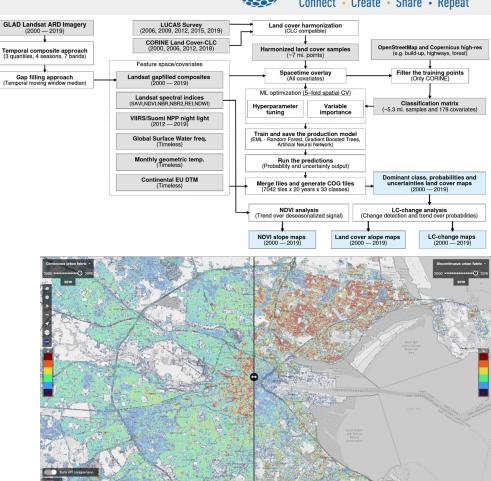
200+ covariates

43 class probabilities

Prediction uncertainty

Data viewer

Open data, open source code



https://doi.org/10.21203/rs.3.rs-561383/v4

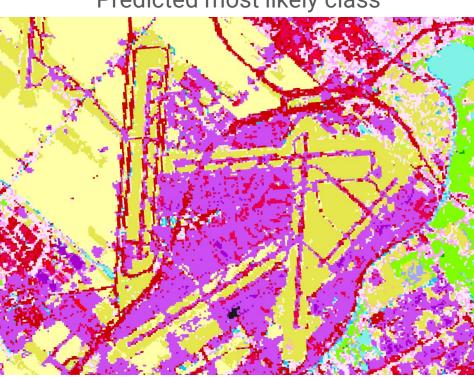
Example of limits: Schiphol Airport, NL



GLAD Landsat ARD - RGB

Predicted most likely class



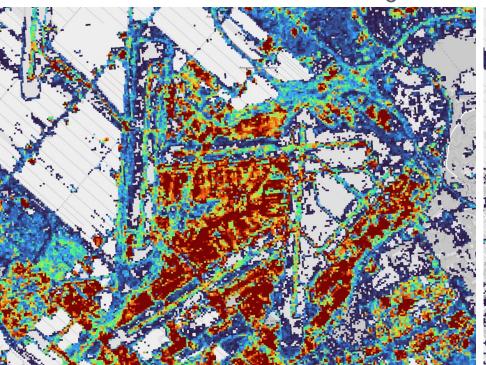


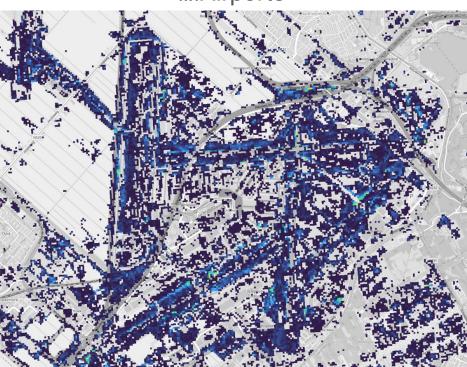
Schiphol predicted probabilities



Industrial & Commercial buildings

....Airports





Incorporating spatial context with CNNs



Convolutional neural networks

Very successful in computer vision & remote sensing

1,	1,	0	0
1,	1 _{×0}	1	0
0,×0	1,	1	1
0	1	1	0
1	1	0	0
	1 _{x0} 0 _{x0} 0	$\begin{array}{c c} 1_{x_0} & 1_{x_1} \\ 1_{x_1} & 1_{x_0} \\ 0_{x_0} & 1_{x_1} \\ 0 & 1 \\ 1 & 1 \end{array}$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$

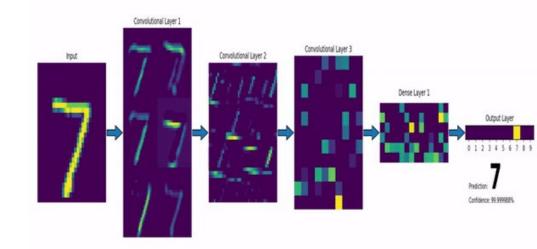


Image

Convolved Feature

Training data limitations:

- Need large amounts*
- Need seamless labels*
- Need labels at pixel level*



LUISA Basemap



46 Classes:

11 Agriculture

17 Built-up

5 Water

13 Forest/Nature

2 years:

20122018

+-CLC compatible:

Expanded:

- Urban 4 densities + vegetation
- Airport areas + terminals
- Road/rail + stations
- Sport/Leisure built + green

Simplified:

wetlands

High density urban fabric Medium density urban fabric Low density urban fabric Isolated or very low density urban fabric Urban vegetation Industrial or commercial units Road and rail networks and associated land Major stations Port areas Airport areas Airport terminals Mineral extraction sites Dump sites Construction sites Green urban areas Sport and leisure green Sport and leisure built-up Non irrigated arable land Permanently irrigated land Rice fields Vineyards Fruit trees and berry plantations Olive groves **Pastures** Annual crops associated with permanent crop Complex cultivation patterns Land principally occupied by agriculture Agro-forestry areas Broad-leaved forest Coniferous forest Mixed forest Natural grassland Moors and heathland Sclerophyllous vegetation Transitional woodland shrub Beaches, dunes and sand plains Bare rock Sparsely vegetated areas Burnt areas Glaciers and perpetual snow Wetlands Water courses Water bodies Coastal lagoons

> Estuaries Sea and ocean

Batista, Filipe; Pigaiani, Cristian (2021): LUISA Base Map 2018. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/51858b51-8f27-4006-bf82-53eba35a142c

4 Current work: An escalated pet project



- Train a CNN (UNet) to predict 46 Luisa Basemap classes
- At multiple resolutions, using:
 - Our 30m Landsat 8 ARD data cube
 - A 10m Sentinel-2 ARD data cube
 - A small number of Planet Fusion tiles
- Make EU-wide maps of any year sufficient LUCAS land cover samples
- Validate with LUCAS land cover samples:
 - LUISA Basemaps 2012 & 2018
 - o Predictions at 30m, 10m, and 3m

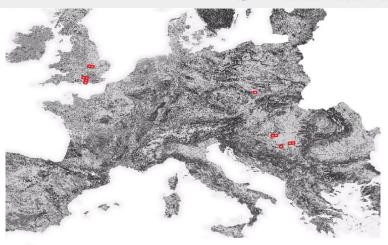
First experiments

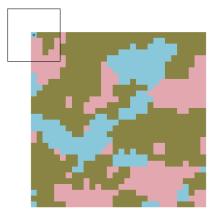


Pilot sites: 12 regions in Europe

Data augmentation:

- Flip
- Mirror
- Moving window





Preliminary results



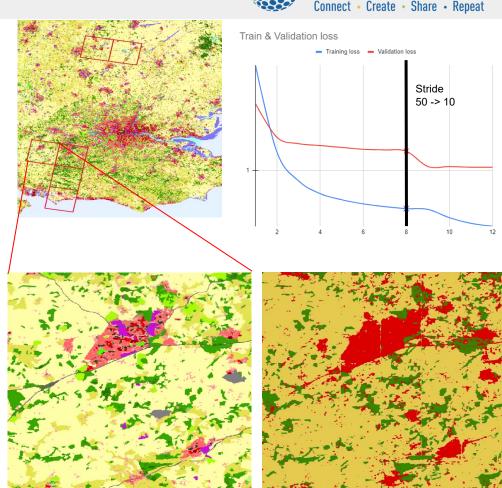
Not enough Planet Fusion data data too sparse & diverse

Preliminary:

- Aggregated to lvl1 (5 classes)
- 30m Landsat ARD
- UK regions

Log loss 1.005 F1 score 0.824

Smaller stride -> less overfitting



Next steps



Design better architecture (multi-input)

Combine multiple resolution data sources

Investigate & Optimize moving window training schedule



Long-term goal



Organize LULC benchmarking dataset

similar to Imagenet (for computer vision)

Y: All 4 levels of LUISA LULC

X: Multiple sources

- 30m Landsat
- 10m Sentinel-2
- 3m Planet

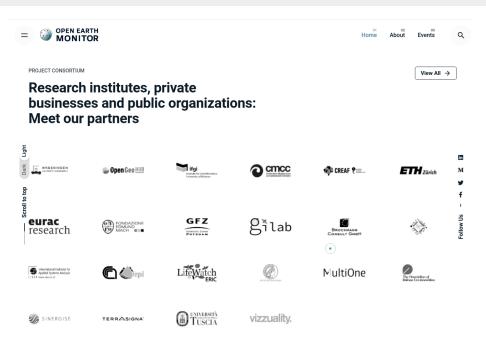
It will be part of project starting now:

OpenEarth Monitor



Open-Earth-Monitor Cyberinfrastructure







Software + EO Data + Community

Martijn Witjes



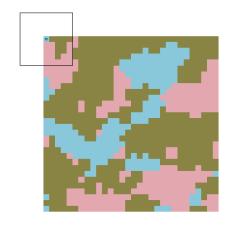
martijn.witjes@opengeohub.org



https://opengeohub.org







Questions

