Large-scale Mapping of Arctic Coastal Infrastructure using Copernicus Sentinel Data and Machine Learning and Deep Learning Methods

Pointner, G.; Bartsch, A.; Ingeman-Nielsen, T.

Link to article, DOI:
10.5281/zenodo.3826154

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Large-scale Mapping of Arctic Coastal Infrastructure using Copernicus Sentinel Data and Machine Learning and Deep Learning Methods

G. Pointner[1,2], A. Bartsch[1,2], T. Ingeman-Nielsen[3]

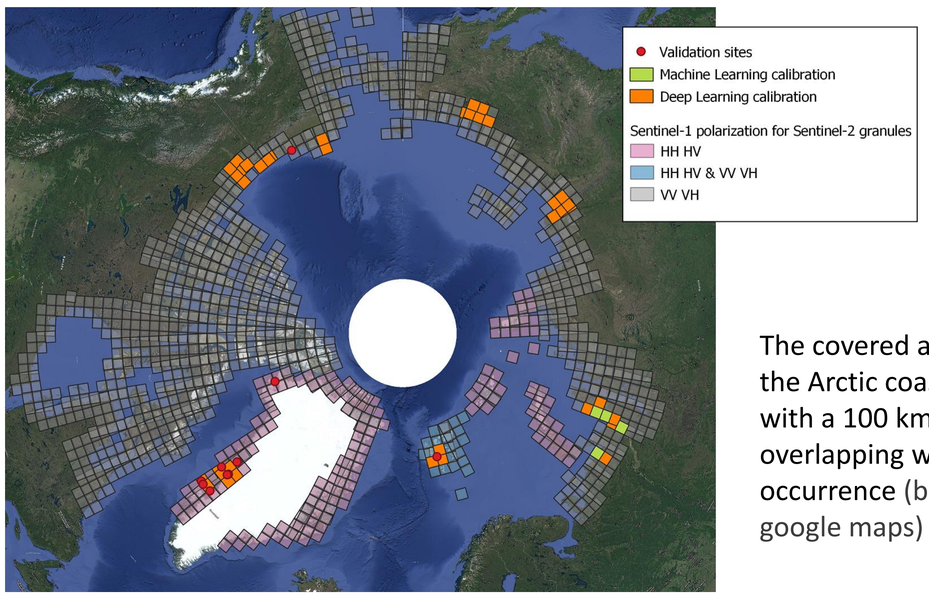
(1) b.geos, Korneuburg, Austria; (2) Austrian Polar Research Institute, Austria; (3) Technical University Denmark

This project is funded by the European Union

Introduction

- The climate change induced increased warming of the Arctic is leading to an accelerated thawing of permafrost, which can cause ground subsidence. In consequence, buildings and other infrastructure of local settlements are endangered from destabilization and collapsing in many Arctic regions. The increase of the exploitation of Arctic natural resources has led to the establishment of large industrial infrastructures that are at risk likewise. Most of the human activity in the Arctic is located near permafrost coasts. The thawing of coastal permafrost additionally leads to coastal erosion, which makes Arctic coastal settlements even more vulnerable.
- The European Union (EU) Horizon2020 project "Nunataryuk" aims to assess the impacts of thawing land, coast and subsea permafrost on the climate and on local communities in the Arctic. One task of the project is to determine the impacts of permafrost thaw on coastal Arctic infrastructures and to provide appropriate adaptation and mitigation strategies. For that purpose, a circumpolar account of infrastructure is needed.
- We present an automated workflow for downloading, processing and classifying Sentinel-2 (optical) and Sentinel-1 (Synthetic Aperture Radar) data in order to map coastal infrastructure with circum-Arctic extent, developed on a highly performant virtual machine (VM) provided by the Copernicus Research and User Support (RUS). We further assess the first classification results mapped with two different methods, one being a pixel-based classification using a Gradient Boosting Machine and the other being a windowed semantic segmentation approach using the deep-learning framework Keras.

Coverage



The covered area comprises the Arctic coastline with a 100 km buffer overlapping with permafrost occurrence (background: google maps)

- More than 1200 Sentinel-2 granules are included in the analyses. For the majority, Sentinel-1 IW mode acquisitions in VV/VH polarization are available. Greenland and some high Arctic islands are covered in HH/HV or HH only. All combinations are only available over Svalbard.
- Calibration sites for both classification approaches are spread across the Arctic. Validation sites include settlements on the Greenland west coast, Longyearbyen and Prudhoe Bay.

Data & Methods

- Sentinel-2 data is a common choice for land cover mapping. Most land cover products only include one class for built-up areas, however. The fusion of optical and Synthetic Aperture Radar (SAR) data for land cover mapping has gained more and more attention over the last years. By combining Sentinel-2 and Sentinel-1 SAR data, the classification of multiple types of infrastructure can be anticipated. Another emerging trend is the application machine learning and deep learning methods for land cover mapping.
- Pre-processing steps:
  - Sentinel-1 IW GRD: border-noise removal, slice-assembly, radiometric calibration, thermal noise removal, terrain correction, incidence angle normalization
  - Sentinel-2 L1C: atmospheric correction, super-resolution, cloud masking, merging of individual scenes, calculation of spectral indices
- Resulting images were stacked and two different approaches are tested for classification with varying polarization combinations:
  - Machine Learning: pixel-based classification using a Gradient Boosting Machine (GBM) from XGBoost (Chen and Guestrin 2016)
  - Deep learning (DL): a windowed semantic segmentation approach using the deep-learning framework Keras (Chollet et al., 2015).
- Calibration data: OpenStreetmap for Deep Learning, Manually collected samples selected with respect to Google Earth imagery for the XGBoost classifier
- Validation data:
  - Prudhoe Bay Cumulative Impact Map, Alaskan North Slope (Raynolds et al. 2014)
  - Dedicated observations within NUNATARYUK for several settlements on Greenland (Ingeman-Nielsen et al. 2018) and for Longyearbyen (Lu et al. 2018) -> Vector geometries thoroughly annotated with information on infrastructure types and materials

Asphalt road and buildings of varying types and roof materials in Ilulissat, Greenland (Ingeman-Nielsen et al. 2018)

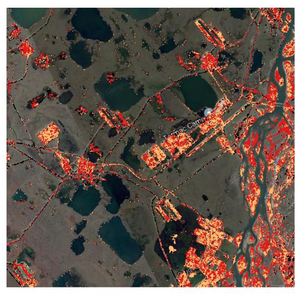
Gravel road and university building with metal roofing in Longyearbyen, Svalbard (Lu et al. 2018)

Preliminary Results

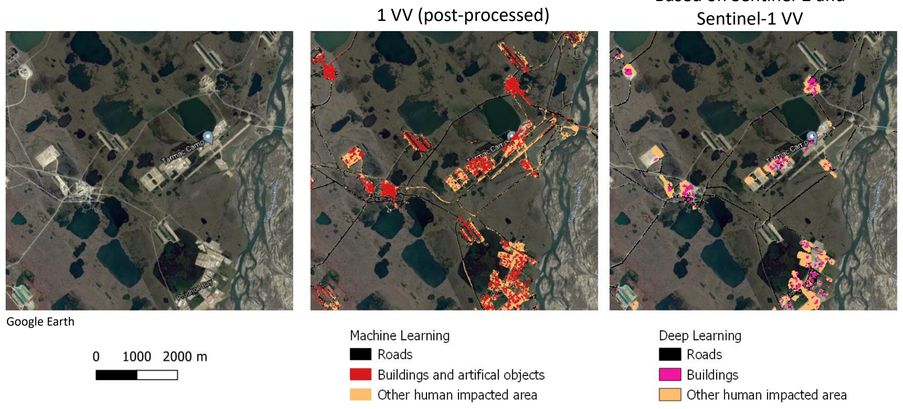
Prudhoe Bay

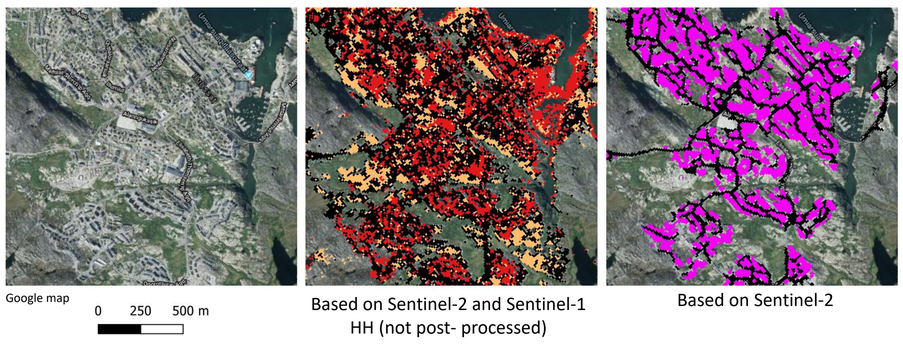
Background mostly tundra, some river beds

only Sentinel-1 VV/VH available



Based on Sentinel-2 only (not post-processed)

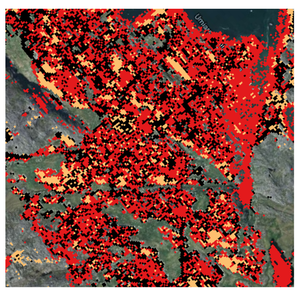




Ilulissat

background mostly bedrock, only patches of vegetation

only Sentinel-1 HH/HV available



Based on Sentinel-2 only (not post- processed)

Dataset references

- Ingeman-Nielsen, Thomas; Vakulenko, Ivan (2018): Calibration and validation data for infratructure mapping, Greenland, link to files. PANGAEA, https://doi.org/10.1594/PANGAEA.895949
- Lu, Wenjun; Aalberg, Arne; Høyland, Knut; Lubbad, Raed; Løset, Sveinung; Ingeman-Nielsen, Thomas (2018): Calibration data for Infratructure mapping in Svalbard, link to files. PANGAEA, https://doi.org/10.1594/PANGAEA.895950
- Raynolds M.K., Walker D.A., Ambrosius K.J., Brown J., Everett K.R., Kanevskiy M., Kofinas G.P., Romanovsky V.E., Shur Y. & Webber P.J. (2014). Cumulative geoecological effects of 62 years of infrastructure and climate change in ice-rich permafrost landscapes, Prudhoe Bay Oilfield, Alaska. Global Change Biology, 20, 1211-1224. http://arcticatlas.geobotany.org/catalog/entries/5470-prudhoe-bay-cumulative-impact-map-a-raynolds-2014

Summary and future work

- Both classification approaches provide useful information but differ strongly in error of omission and error of commission, GBM is considered slightly better in terms of errors of omission, DL better in terms of errors of commission (Bartsch et al., in preparation)
- Both classification approaches require manual post processing in order to provide a useful data product
- A method that incorporates a combination of both approaches may be able two make use of the respective advantages of the presented methods
- Detailed validation datasets will enable assessment of classification results with respect to type and material of infrastructure objects

References:

Widhalm, B., Bartsch, A., and Goler, R. (2018). Simplified normalization of C-band synthetic aperture radar data for terrestrial applications in high latitude environments. Remote Sensing. doi:10.3390/727rs10040551
Bartsch, A., Widhalm, B., Leibman, M., Ermokhina, K., Kumpula, T., Skarin, A., et al. (2020). Feasibility of tundra vegetation height retrieval from Sentinel-1 and Sentinel-2 data. Remote Sensing of Environment 237, 111515. doi:10.1016/j.rse.2019.111515
Tianqi Chen and Carlos Guestrin (2016): XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
Francois Chollet (2018): Deep learning with Python. Manning Publications, Shelter Island. ISBN 9781617294433; Chollet, F., et al. (2015). Keras. https://keras.io.
Palazzo, F., Šmejkalová, T., Castro-Gomez, M., Rémondière, S., Scarda, B., Bonneval, B., ... & Mora, B. (2018). RUS: A New Expert Service for Sentinel Users. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 2, No. 7, p. 369).
Annett Bartsch, Georg Pointner and Thomas Ingeman-Nielsen (in preparation). Towards circumpolar mapping of Arctic settlements and infrastructure based on Sentinel-1 and Sentinel-2

Nunataryuk: https://nunataryuk.org/