A multi-sensor approach for characterising human-made structures by estimating area, volume and population based on sentinel data and deep learning

Fibæk, Casper Samsø; Keßler, Carsten; Arsanjani, Jamal Jokar

Published in:
International Journal of Applied Earth Observation and Geoinformation

DOI (link to publication from Publisher):
10.1016/j.jag.2021.102628

Publication date:
2021

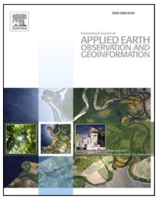
Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Fibæk, C. S., Keßler, C., & Arsanjani, J. J. (2021). A multi-sensor approach for characterising human-made structures by estimating area, volume and population based on sentinel data and deep learning. *International Journal of Applied Earth Observation and Geoinformation*, *105*, Article 102628. https://doi.org/10.1016/j.jag.2021.102628

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



A multi-sensor approach for characterising human-made structures by estimating area, volume and population based on sentinel data and deep learning

Casper Samsø Fibæk [a,b,*], Carsten Keßler [a,c], Jamal Jokar Arsanjani [a]

[a] *Department of Planning, Aalborg University, Copenhagen, A.C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark*
[b] *NIRAS Consulting, Sortemosevej 19, 3450 Allerød, Denmark*
[c] *Bochum University of Applied Sciences, Am Hochschulcampus 1, 44801 Bochum, Germany*

ARTICLE INFO

Keywords:
Earth observation
Copernicus programme
Deep learning
Structures
Population

ABSTRACT

Several global and regional efforts have been undertaken to map human-made settlements and their characteristics, including building material, area, volume, and population. However, given the unprecedented amount of Earth observation data and processing power available, there is a timely need for developing novel approaches for mapping these characteristics at higher spatial and temporal resolution. Such information is key to effectively answering questions related to population growth, pollution, disaster management, risk assessments, spatial planning, and even generating business cases in *peri*-urban and rural areas. While such data is available from mapping agencies or commercial companies in some countries, there are many countries where this is not the case. The main objective of this study is to propose an Inception-ResNet inspired deep learning approach to estimate the characteristics and location of human-made structures, including estimates of population, based on Earth observation data from the Copernicus Programme. The study investigates the effects on prediction accuracy using data from different orbital directions and interferometric coherence from Sentinel 1 data and different band combinations of Sentinel 2 data as model input variables. The model is trained and evaluated on a nationwide Danish case study, where the national mapping agency provides high-quality open data on human-made structures, which serves as the ground truth data for the study. Our findings reveal that it is possible to design models that, on average, perform within 2.6% total absolute percentage error for area predictions, 7.7% for volume and 17% for population at 10 by 10 m scale using only Copernicus data and deep learning models. The models achieved 98.68% binary accuracy for extracting structural area when all test sites were merged. Combining Sentinel 1 and 2 input variables yielded the best results, while adding interferometric coherence did not significantly improve accuracy. Furthermore, including data from both orbital directions of the Sentinel 1 constellation significantly improved model performance.

1. Introduction

Monitoring and planning for urbanisation, estimating the local population in intercensal years, improving financial inclusion by mapping under-served communities, and organising disaster responses are examples of endeavours that benefit from reliable data on the area, volume, population and location of building infrastructure (Fibæk et al., 2021). However, access to quality data varies significantly across the globe. Some countries provide structural data from LIDAR surveys, large scale national topographic mapping projects, and detailed building registries. A similar amount of quality data are not available in many countries, especially data-poor countries in the Global South (Woon 2013). Moreover, these countries often face complex planning issues, including significant urban expansion, which is projected to continue for the next decades (Jiang and O'Neill 2017). This study shows that it is possible to generate estimates of structural characteristics using data from the Copernicus programme. The novelty of this study comes from the temporal and spatial resolution made possible by the proposed methodology, the sole reliance on data from the Sentinel satellites, and predictions of area, volume and population at the individual structure scale.

A model to map the characteristics of human-made structures that

[*] Corresponding author at: Casper Samsø Fibæk, Lyneborggade 8, 2300 Copenhagen S, Denmark.
E-mail addresses: cfi@plan.aau.dk, cfi@niras.dk (C.S. Fibæk), kessler@plan.aau.dk (C. Keßler), jja@plan.aau.dk (J.J. Arsanjani).

https://doi.org/10.1016/j.jag.2021.102628
Received 9 September 2021; Received in revised form 14 November 2021; Accepted 15 November 2021
Available online 27 November 2021

use open data policy satellites can lower costs and barriers to continuous monitoring of structures and populations. For countries with existing data infrastructure on structures, an Earth Observation-based model could increase the temporal resolution of existing datasets. The high temporal resolution of the Sentinel satellites can make large-scale, high-resolution mapping more affordable by enabling targeted topographic updates. Synthetic Aperture Radar (SAR) data from the Sentinel 1 satellite make models more resistant to weather effects in time-critical scenarios and enable mapping in areas with perpetual cloud cover (Khabbazan et al. 2019). Mapping the characteristics of buildings helps describe and estimate dynamics in cities, such as greenhouse gas emissions, urban heat islands, storage inventory, floor area, and energy consumption (Frantz et al. 2021).

Earth observation and deep transfer learning can help alleviate the impact of the global data disparity (Jasper et al. 2021). However, a great amount of ground truth data is usually needed to train the base models (Herfort et al. 2021). This study, therefore, proposes training deep learning models on the data available in Denmark, both to show the feasibility of the approach and to enable future deep transfer learning to other geographical regions for mapping building characteristics. The models developed can be used as an input layer in other related geospatial tasks, such as classifying urban settlement types or structure types. Combining structural information with census data and land cover classification makes it possible to create population density estimates for places where ground truth, similar to the Danish population ground truth dataset, is not available (Leasure et al. 2020; Tatem 2017).

Using Denmark as a testing location ensures a controlled environment where complete ground truth data coverage is available. The location allows a focus on the design of the model and input variables. The models are beneficial for Denmark to conduct targeted topographic updates, extract historical and current structures, and assess local population density. The proposed methodology's predictions at 10 by 10 m resolution are lower than what is achievable using satellite imagery with submeter spatial resolution. There are, however, significant benefits to approaches based on data from the Copernicus programme: No satellite tasking is required, high temporal resolution is available for most of the globe, it is free of charge, and it requires less processing power and storage capabilities than higher-resolution methods. The combination of active and passive sensors and the use of a broad range of the electromagnetic spectrum ensures high robustness in the predictions versus methods based solely on passive sensors capturing the visible spectrum.

Data from the Sentinel satellites are well suited to form the basis of these models. The Sentinel 1 constellation provides C-Band SAR data, and the Sentinel 2 constellation provides high-resolution multi-spectral data. Together, they have been proven highly effective at mapping and classifying urban clusters and settlement types (Qiu et al. 2020; Semenzato et al. 2020).

Hence, this study aims to:

- Investigate the possibility of mapping structures and their characteristics at a 10 by 10 m spatial resolution using publicly available satellite data.
- Investigate the effect of using different input variables on a multi-sensor approach using an Inception-ResNet inspired Convolutional Neural Network and determine the effects of using different tile sizes on the models' accuracy.

The remainder of this paper is structured as follows: Section 2 presents related work to the study and background context. Section 3 describes the study area, followed by Section 4, depicting the methodology behind the analysis. Section 5 presents the achieved results, while Section 6 and 7 present the discussion and conclusions of the paper.

2. Related work

This study investigates the use of the Sentinel 1 and 2 satellites to create datasets on structural characteristics – their area, volume, and human population. Weijia Li et al. (2019) and Microsoft (2021) showed examples of extracting building footprints from very high-resolution imagery, either from satellites or aerial imagery. The two approaches presented are different from the one applied in this study, in both the sensors chosen and the geospatial scale, but the goal of mapping the location and characteristics of structures is similar. The models applied in the two studies were trained on data from the visible spectrum, whereas this study uses all the 10 and 20 m bands available from Sentinel 2 satellites, which includes Near Infra-Red (NIR) and Short Wave Infra-Red (SWIR) bands, as well as both backscatter and coherence from the Sentinel 1 satellite in VV and VH polarisations. Weijia Li et al. (2019) and Microsoft (2021) produced vectorised building footprints, which is not done in this study, as the Sentinel data is of a spatially coarser resolution. Through their programme "AI for Humanitarian Action" (Microsoft 2021), Microsoft has published building footprints generated for Uganda and Tanzania built upon the deep learning model specified in Tan and Le (2019). The project was conducted in collaboration with the Humanitarian OpenStreetMap Team. The data provided by the project can be used as part of the ground truth dataset for the models' proposed here. Google has released building footprints for 64% of Africa built upon modified U-Net and ResNet models and 50 cm satellite imagery using the visible spectrum (Sirko et al. 2021). The general approach described in Sirko et al. (2021) is similar to Weijia Li et al. (2019) and Microsoft (2021). However, they introduced noisy student learning iterations to improve their results further (Xie et al. 2020). This study focuses on a multi-sensor approach for data of a significantly spatially coarser character than Microsoft (2021) and Sirko et al. (2021) while using the additional information offered by expanding the breadth of the utilised parts of the electromagnetic spectrum and combining passive and active sensors.

Gao and Cui (2020) suggested using transfer learning to lessen healthcare disparity, which arises from the data inequality between ethnic groups. This study proposes that a similar approach could apply to global mapping efforts. Up-to-date maps and topographic data are essential for urban planning and disaster management. OpenStreetMap is a data source heavily used in the humanitarian sector, but data inequality is also present here. While 28 % of the mapped buildings are located in low and medium development index countries, the countries are home to 46 % of the global population (Herfort et al. 2021). Deep learning models trained on areas with high data availability can be applied to lessen the data inequality.

Frantz et al. (2021) investigated the use of Sentinel 1 and 2 data time series to map the building heights of structures in Germany using a Support Vector Machine (SVM) regression model. Their general approach is similar to the proposed method for predicting structural volume. However, this study relies on deep learning, terrain models, and public structure footprints instead of 3D building datasets and SVM regression. The focus of this study is not establishing the building height but the structural area, volume, and population on a given 10 by 10 m pixel. Frantz et al. (2021) suggested using both orbital directions for Sentinel 1 to lessen the effect of building orientation, and this finding is incorporated into the methodologies presented in this study. Xuecao Li et al. (2020) investigated a problem of estimating building heights using single orbital direction Sentinel 1 over the United States at the 500 by 500 m scale and saw good results, which along with Frantz et al. (2021), show that using sentinel imagery to estimate building volume is feasible.

Xinghua Li et al. (2021) used a modified U-Net model called U-Net++ (Zhou et al. 2018) and a process called Deep Translation based Change Detection Network, which combines SAR and optical imagery into the same feature domain before feeding it to a change detection network. This study also investigates how to combine the active and passive sensors, taking a different approach by using the inception module approaches presented in Szegedy et al. (2017) and designing a neural network to incorporate the data from different sensors in different branches and merging them before the upsampling blocks.

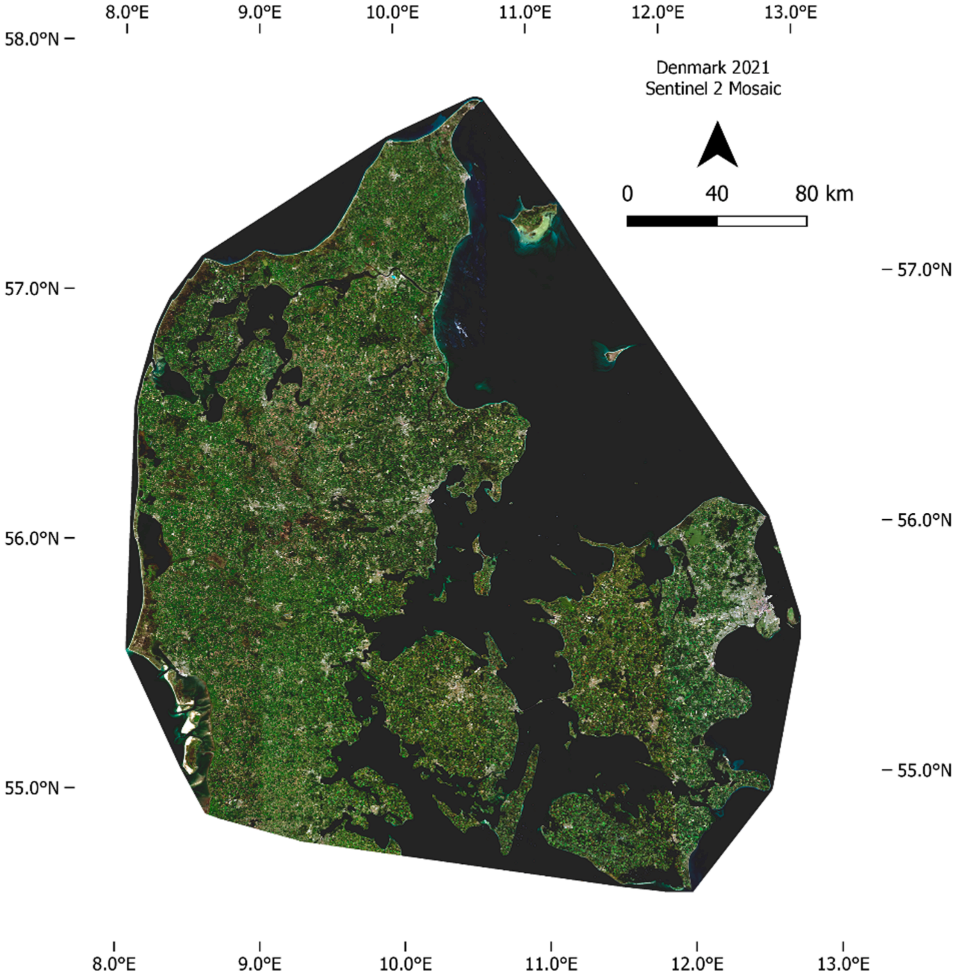


Fig. 1. Denmark, without the Island of Bornholm. Sentinel 2 Mosaic from Spring 2020 to early Summer 2020.

Global maps of population density and building footprints exist, such as WorldPop and Ecopia (Ecopia 2021; Esch et al. 2017; Tatem 2017). WorldPop combines multiple data sources, such as night lights, health facilities and Landsat imagery, into their population estimates made using a random forest model presented in Stevens et al. (2015). This study contributes to the research gap by using only data from the Copernicus programme and deep learning approaches to map several structural characteristics through regression models.

3. Study area

The study area is the whole of Denmark shown in Fig. 1, excluding the island of Bornholm in the Baltic Sea, which serves as one of the testing areas excluded from the training material. Denmark contains urban clusters, heathland, coastal areas, large scale agriculture and lakes. The country contains several cities with large industrial zones, dense residential neighbourhoods, holiday homes, and extensive suburbs. Denmark is home to 5.8 million people, and there are 5.7 million registered buildings. The land area of Denmark is 42,933 km^2, and all of it is included in the study area, either for testing or training.

Most of Denmark has been recently scanned in a national update of the LiDAR-based national terrain model. The availability of recently updated data means that it is possible to do up-to-date structural volume calculations for most of the country.

The major cities in the country are predominantly located along the coastal regions and fjords. The buildings are mostly one to two-floor detached housing in suburban areas with clusters of 6–8 floor buildings in the urban centres. The tallest building in Denmark is Herlev Hospital at 120 m in the Copenhagen metropolitan area. Frequent cloud-free Sentinel 2 images are sparse outside of the summer months (Danish Meteorological Institute 2019).

Table 1
Subset areas used for testing the models.

Test Area	Type	Size km^2	Population
Holstebro	Mixed	216.44	41,991
Aarhus	Urban	103.29	140,454
Samsoe	Rural	344.26	3456
Odense	Urban	298.37	203,377
Bornholm	Rural	588.77	39,523

Five test areas were excluded from the training material and used to test the accuracy of the developed models. Table 1 and Fig. 2 show the test areas.

The five test areas chosen contain a mix of different structure types, from the urban areas of Aarhus and Odense to the mixed urban/agricultural area of Holstebro and the rural islands of Bornholm and Samsoe, which are popular vacation destinations home to many vacation homes and campsites.

4. Methodology

Three different models were trained to predict three levels of characteristics about human-made structures: Area, volume, and the number of inhabitants. Separate ground truth datasets were created for each model, and Sentinel 1 and 2 data serve as the input variables. Table 2

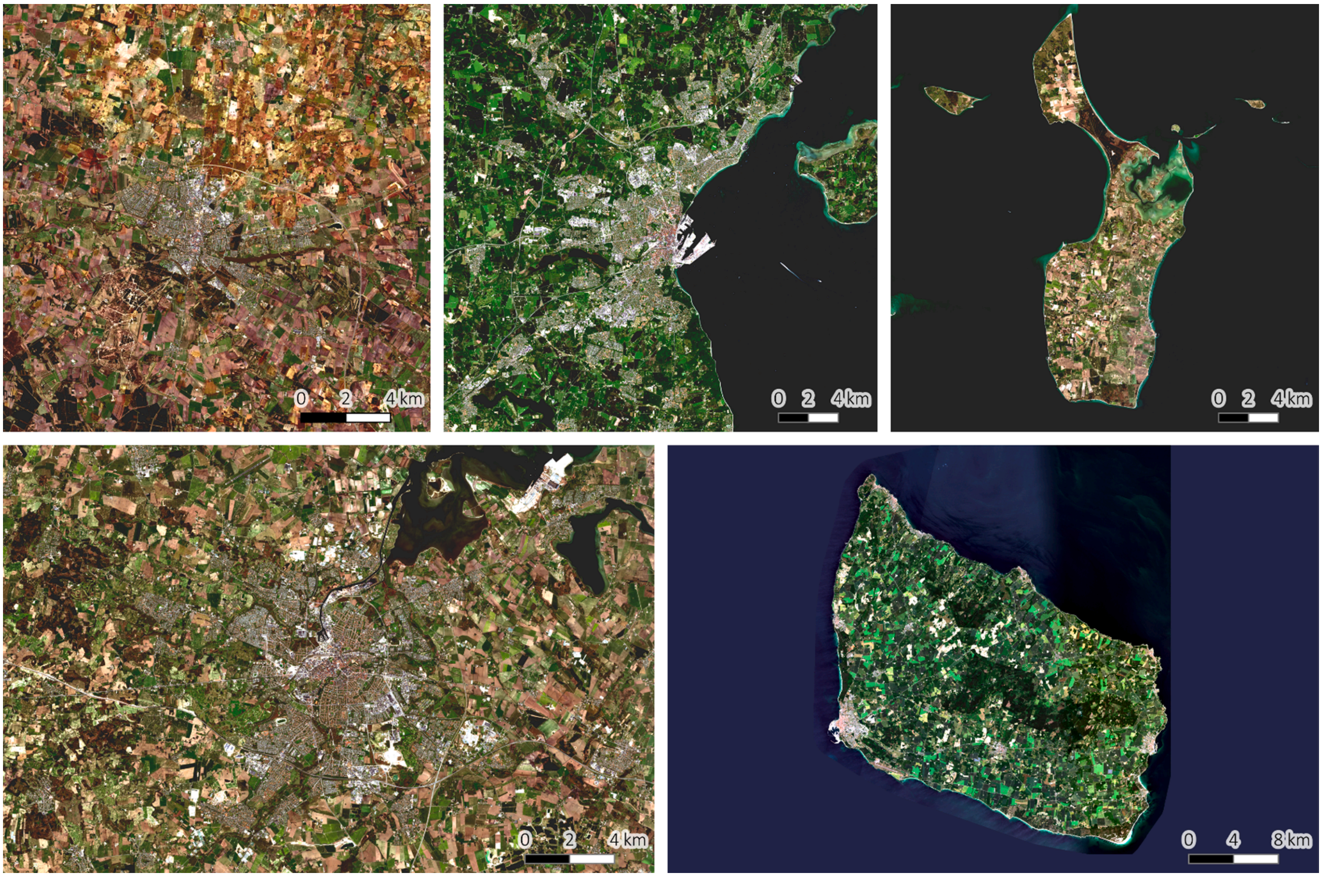


Fig. 2. The areas used for testing, Sentinel 2, 2021. From the top left: Holstebro, Aarhus, Samsoe Island, Odense and Bornholm Island. Not all of Aarhus Municipality was part of the testing area.

Table 2
Overview of data sources used.

Dataset	Source	Purpose
Sentinel 1 GRD and SLC	ESA SciHub	Training
Sentinel 2 Level 2A	ESA SciHub	Training
GeoDenmark Buildings	Danish Map Supply	Labels
GeoDenmark Surface Model	Danish Map Supply	Labels
GeoDenmark Terrain Model	Danish Map Supply	Labels
Denmark Addresses WebAPI	DAWA	Labels
Danish Population Data	Danish Statistics	Labels
Danish Administrative Boundaries	DAGI	Labels

shows the different data sources used to generate the ground truth and input variables.

4.1. *Ground truth*

The datasets that serve as the ground truth have increasing conceptual complexity. First is the sum of the area (2D) of structures intersecting a given 10 by 10 m pixel, second is the volume (3D), and third is the number of humans living in structures within each pixel. An example of the ground truth data used in this study is shown in Fig. 3.

4.1.1. *Area*

Data from GeoDanmark – a collaboration between the Danish Ministry of Data Supply and Efficiency and the 98 Danish municipalities – provides the ground truth for building footprints. GeoDenmark is responsible for maintaining up-to-date and accurate standardised geospatial vector data of Denmark (GeoDenmark 2021). The used footprints do not contain windmills, bridges and other large infrastructure elements, and the accuracy of the building vector dataset is considered very high (Flatman et al. 2016). Both OpenStreetMap and GeoDenmark have comparable datasets available for buildings in Denmark, as shown in Table 3.

GeoDenmark was chosen due to the higher coverage. In many countries, OpenStreetMap will be the best provider of free and up to date on buildings (Panek and Netek 2019).

Denmark was divided into a grid of 10 by 10 km tiles to distribute the processing. The grid was aligned with the Sentinel 1 and 2 mosaics before rasterisation. The building vectors were rasterised to 40 cm resolution with the value one for buildings and zero otherwise. Each tile was then resampled using the sum to 10 by 10 m resolution. The result is an aligned raster with values ranging from 0 to 100 corresponding to the structural area present on each pixel.

4.1.2. *Volume*

The volume of structures on each pixel was calculated using the national LiDAR-derived surface and terrain models in their native 40 cm resolution and the GeoDenmark building dataset. The digital elevation models of Denmark are updated on a rolling basis, with data in some cases being up to five years old. The temporal difference between the building vector layers, which are updated annually, and the elevation models, means that volume calculations for some municipalities are significantly out of sync. The errors are primarily due to the construction of new structures, captured in the GeoDenmark building dataset but not yet captured by the elevation models. An example of the two datasets being out of sync is shown in Fig. 4. Buildings captured by the terrain models but not the GeoDenmark dataset were not validated.

The volume of the structures was calculated using the difference between the surface and the terrain model rasterised and clipped to the

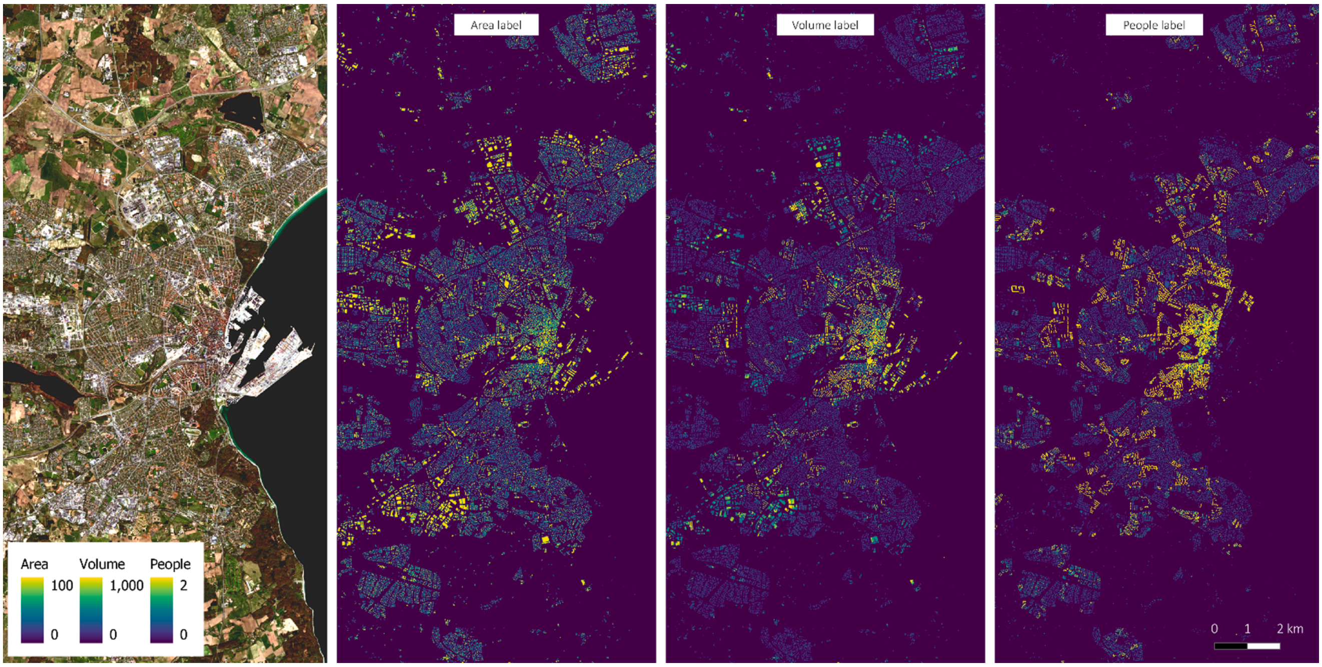


Fig. 3. Ground truth used in the study shown for the urban area of Aarhus.

Table 3
Building data provider comparison for Denmark. Data collected August 2021.

Dataset	Source	Buildings	Covering Area	License
GeoDenmark	Danish Map Supply	5.69 million	738.52 km^2	Free
OpenStreetMap	GeoFabrik	3.51 million	690.69 km^2	Free & Open

GeoDenmark building footprints. The resulting layer was then resampled and aligned to match the Sentinel imagery as the area layer. In some cases, as seen in Fig. 4, the footprints were not in sync with the terrain models or trees partially covered the structure.

A multilayer perceptron artificial neural network was trained to predict the volume of out of sync buildings based on the area, perimeter, and the isoperimetric quotient measure (IPQ) of compactness (Li et al., 2013).

$$IPQ = \frac{4^{*}\pi^{*}area}{perimeter^{2}}$$

Buildings that were not in sync with the terrain models are defined in this study as “Spatially dissolved structures with an area to volume ratio of less than one and area above one.” 7.5 per cent of the buildings matched this criterion and were selected for volume estimation using the neural network model. The data was split into training, test, and validation sets to train the neural network. The model accuracy assessment used five stratified shuffle splits of 50 % of the dataset with a 10 % test set. The results are shown in Table 4. The resulting estimates from the fully trained network were used in the rest of the study as part of the ground truth for volume estimations.

The estimates for building volumes enabled the use of the entire building footprint dataset from GeoDenmark.

4.1.3. Population

The number of people living in structures within each pixel was estimated using a combination of four datasets: The GeoDenmark buildings dataset, the Danish parish boundaries, the Danish address registry, and population statistics from the National Statistics Office. The population statistics come at the parish level, which was then spatially joined to the parish boundaries. All addresses in Denmark are geocoded and generally located within a building. A building that houses multiple families has multiple addresses. All the buildings were dissolved and filtered based on the intersection with a valid address point. The population of each parish was distributed equally amongst the access point addresses located within the parish and the intersecting building. The population per building was then distributed amongst each covered pixel and resampled, similar to the area and volume ground truth process.

The population dataset combines daytime and nighttime populations, as businesses and organisations’ addresses are also included (Qi et al. 2015). In most cases, non-residential buildings will only have one “access point address” attached and therefore get a minimal share of the distributed population. If the spatial extent of the analysis is larger than the parish level, the sum of the population within each parish will always correspond to the National Statistics Office numbers.

4.2. Sentinel 1

Sentinel 1 imagery has been proven useful for urban land cover classification (Abdikan et al. 2016). SAR is subject to a double bounce effect in sensing scenarios where walls face the sensor. Utilising the double bounce effect of SAR data in urban areas is effective for mapping structures, and Semenzato et al. (2020) and Frantz et al. (2021) have shown that Sentinel 1 can be used for estimating the height of structures. Over some areas of the globe, Sentinel 1 data is only available from one orbital direction (European Space Agency 2015). This study tests the impact of having access to only one orbital direction on the accuracy of the model estimates.

All available Ground Range Detected (GRD) data over Denmark was collected from spring to early summer 2020, and 2021 and Single Look Complex (SLC) data was collected from 15 March to 31 March 2020 over the Central Denmark Region. The data were preprocessed using the Graph Processing command-line Tool (GPT), part of the SNAP toolbox (Agency European Space 2021). The GRD files were processed using a standard GRD preprocessing toolchain, as illustrated in Fig. 5.

Signals from different orbital directions are impacted by different corner reflectors. Fig. 6 shows the difference between the two orbital

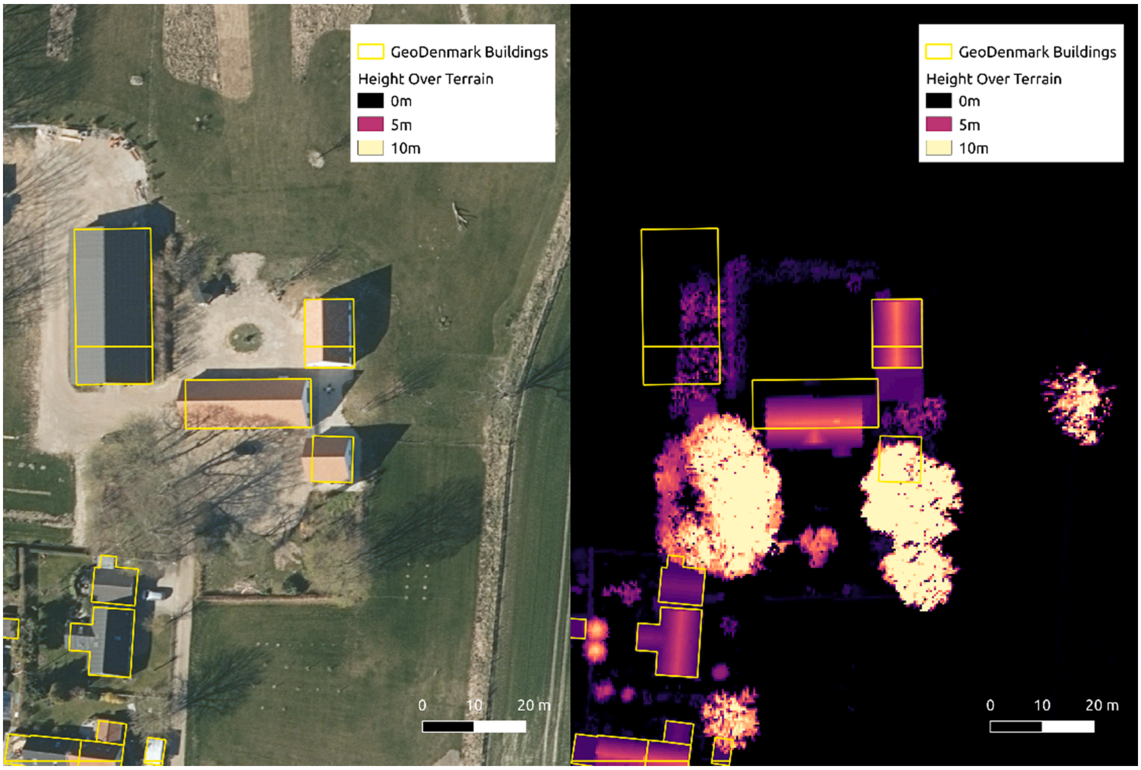


Fig. 4. Left: GeoDenmark Orthophoto and Buildings, Right: GeoDenmark buildings and the height over terrain.

Table 4
Results from the volume estimation of structures using their area, perimeter, and circularity.

Metric	Result	1 σ
Mean Absolute Error (MAE)	128.7 m^3	1.4 m^3
Mean Absolute Percentage Error (MAPE)	25.1 %	0.4 %
Median Absolute Error (MeAE)	29.3 m^3	0.6 m^3
Median Absolute Percentage Error (MeAPE)	17.8 %	0.1 %

directions on backscatter intensity in the ascending and descending direction and the difference and average between them. Notice the example of the significant differences in the Northeast of Aarhus highlighted with a white circle.

Interferometric coherence between Sentinel 1 SLC data pairs was calculated as defined in Jacob et al. (2020) described in Fig. 7 and visualised in Fig. 8. Interferometric coherence generally decreases faster in vegetated or flooded areas than in urban built-up areas (Koppel et al. 2015).

Processing SLC data is a data-heavy and time-consuming process compared to analysing the level 2 GRD data. Only the VV polarisation was used for the coherence calculations. All of the SLC images were matched with their closest baseline.

After processing coherence and backscatter data for the downloaded scenes, the data were multi temporally filtered using the median values in a 3D ellipsoidal temporal distance weighted kernel with the images sorted by date for overlapping points to reduce the speckle noise in the final imagery (Trouvé et al. 2003) and then processed to decibels.

4.3. Sentinel 2

Multi-spectral atmospherically corrected imagery from Sentinel 2 was collected for Denmark captured during spring to early summer 2020 and 2021. The project area of Denmark covers 17 Sentinel 2 100 by 100 km tiles, and 156 scenes were used to generate the two mosaics.

The Sentinel 2 mosaics were created using a custom Sentinel 2 processing toolchain, which uses a novel methodology for generating mosaics inspired by Sen2Mosaic (Bowers 2021) and Sen2Three (Mueller-Wilm 2021). The toolchain evaluates the quality of each tile and its pixels and creates a mosaic composed of as few source images as possible while fulfilling a predetermined quality threshold. Spectral harmonisation is done through MAD-matching (Leys et al. 2013).

4.4. Data preprocessing

The input imagery was cut into tiles as input variables for the CNN architecture. The following tile sizes in 10 by 10 m pixels were investigated: 16 × 16, 32 × 32, 64 × 64, and 128 × 128. For each tile, the 20 m resolution Sentinel 2 bands were tiled at half the size of the 10 m inputs. The patches were extracted with overlaps offset at half the tile

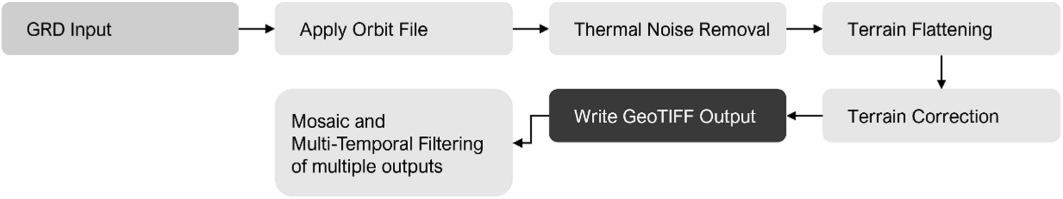


Fig. 5. Sentinel 1 GRD Preprocessing toolchain.

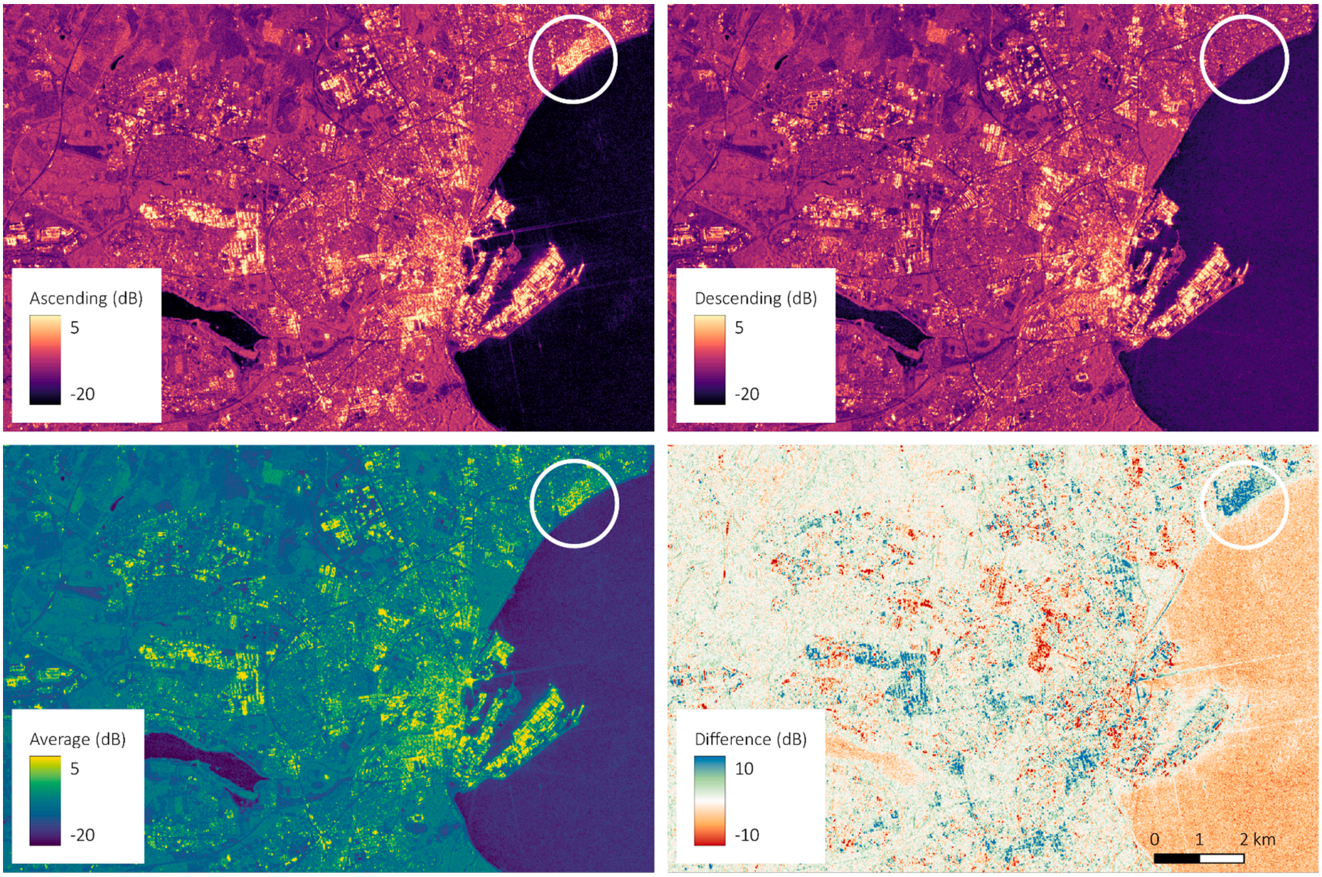


Fig. 6. Sentinel – Effect of Orbital Direction on the backscatter intensity (Aarhus, Sentinel 1). Significant difference highlighted with a circle.

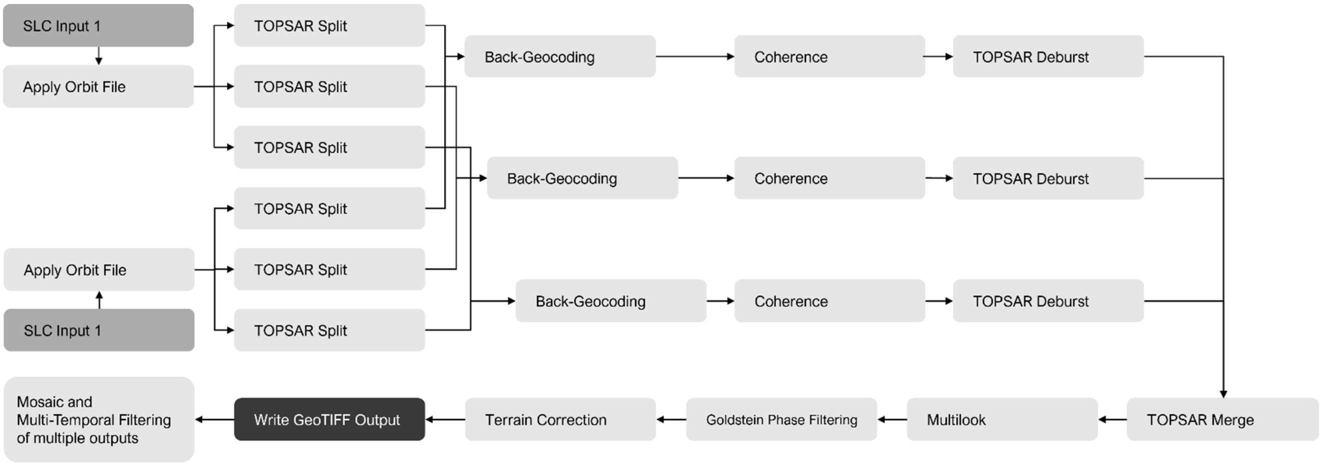


Fig. 7. SLC Preprocessing toolchain.

resolution; that is an additional 8 × 8 offset for the 16 × 16 patches, as shown in Table 5. The larger tile sizes could reduce the amount of border noise in the joined rasters but come at a much larger memory footprint – meaning smaller batch sizes are required to train the model. Smaller tiles increase flexibility in the model design but are likely to increase border noise.

After patch generation, the datasets were normalised. The Sentinel, 1 backscatter bands, were clipped to between −30 dB and 15 dB and the Sentinel 2 bands between zero and 8000, before they were normalised to values between zero and one.

4.5. Model design

The model works by predicting a combination of physical characteristics and image recognition. The double bounce effect, captured by the Sentinel 1 satellites, and the shadows cast by the structures, captured by the Sentinel 2 satellites, enable the prediction of structural area and volume, while context and texture clues from the imagery allow the model to further qualify the predictions (Frantz et al. 2021). The model architecture is a modified version of the Inception-ResNet architecture described in Szegedy et al. (2017). Rectified Linear Unit (ReLU) was selected as the output activation function as the three ground truth

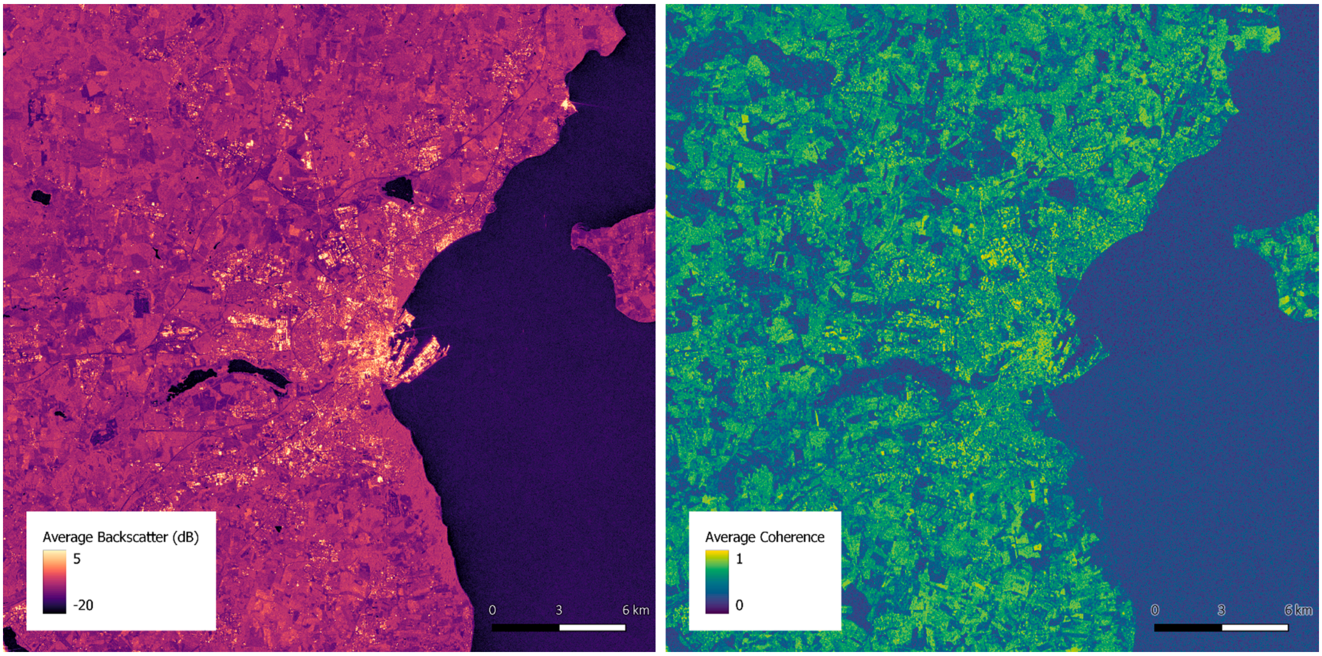


Fig. 8. Backscatter (VV) and Interferometric Coherence (VV) over Aarhus.

Table 5
Example patch sizes and extracted overlaps.

10 m inputs	20 m inputs	Overlap offset 10 m	Overlap offset 20 m
16 × 16	8 × 8	8x, 8y	4x, 4y
32 × 32	16 × 16	16x, 16y	8x, 8y
64 × 64	32 × 32	32x, 32y	16x, 16y
128 × 128	64 × 64	64x, 64y	32x, 32y

Table 6
Model parameters for input variable tests.

Name	Input layers	Parameters
Single	1	3.82 million
Duo	2	4.72 million
Trio	3	5.61 million
Large Models	3	13.39 million

datasets can only be positive and because it has been proven to perform well (Xu et al. 2015). The optimiser uses the lookahead optimiser approach with the Adam optimiser and stepwise learning rate decay (Kingma and Ba 2015; Zhang et al. 2019). The models were fitted three times to different batch sizes while also decaying the learning rate. While the batch sizes were decreased rather than increased, as proposed in Smith et al. (2018), decreasing the batch sizes yielded the best results. The Mean Square Error (MSE) loss function was used for all the models. While other loss functions were explored, MSE consistently produced the smallest total percentage error, which was a key target metric for this study, due to MSE leading to forecasts towards the mean, while MAE leads towards the median (Hyndman and Athanasopoulos 2018).

The predictions are bounded to zero below by ReLU but remain unbounded above. The unbounded top means values of more than 100 is possible for the area of a 10 by 10 m pixel. The unboundedness can be fixed by clipping the output in postprocessing. However, no clipping was done in the comparisons of the predictions.

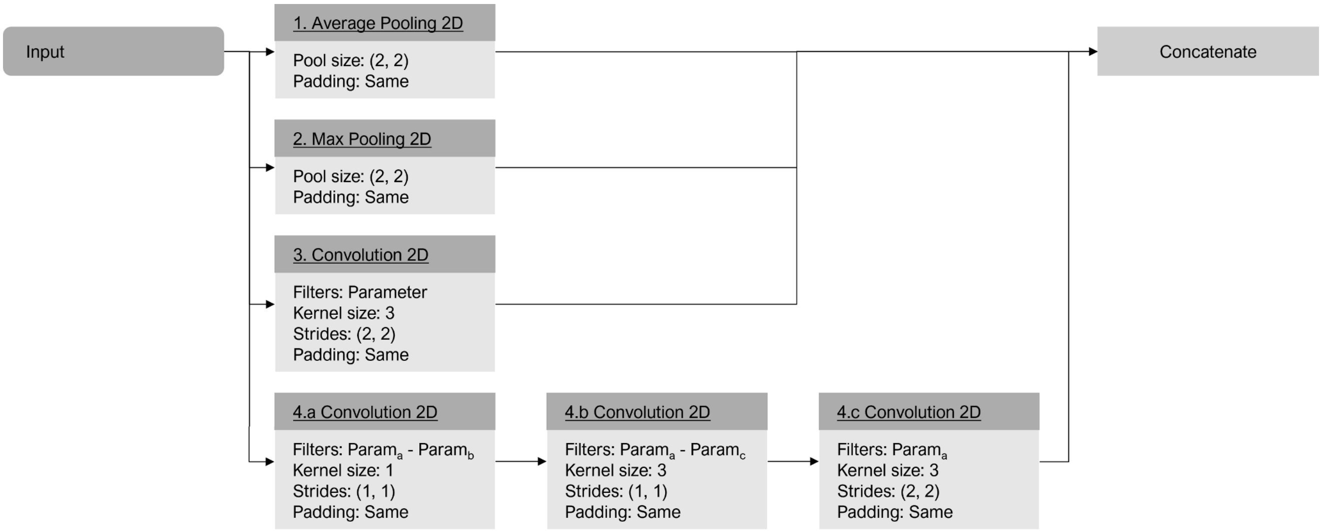


Fig. 9. The Design of the Inception inspired reduction blocks.

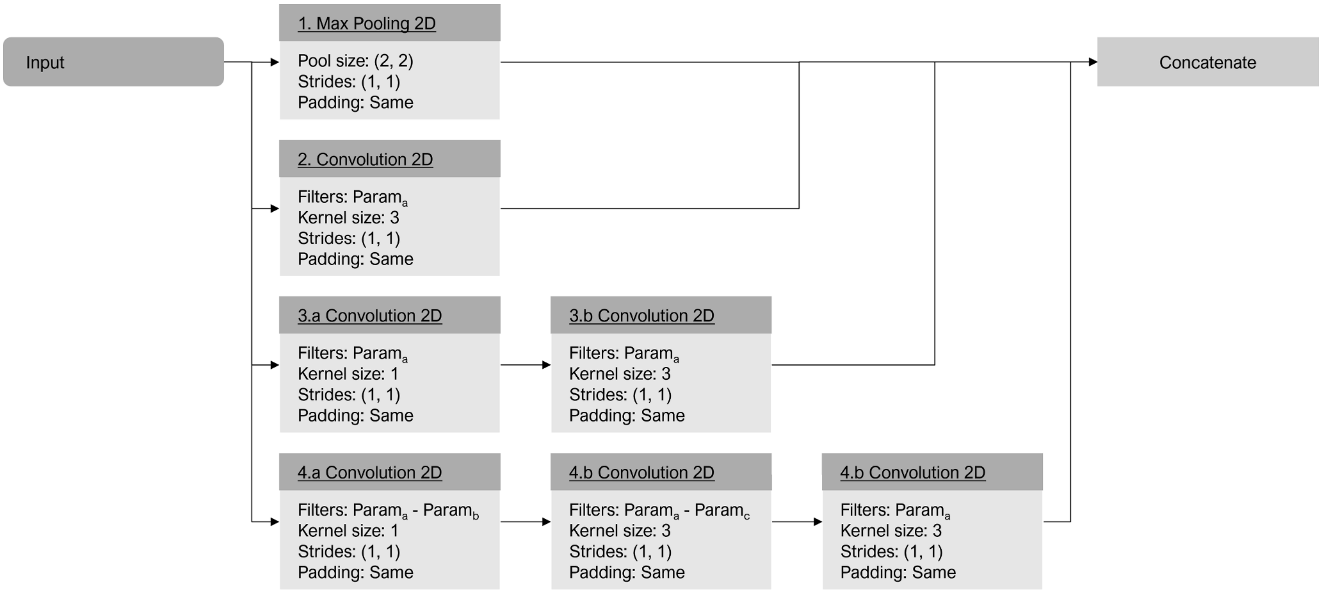


Fig. 10. The Design of the Inception-ResNet inspired convolutional blocks.

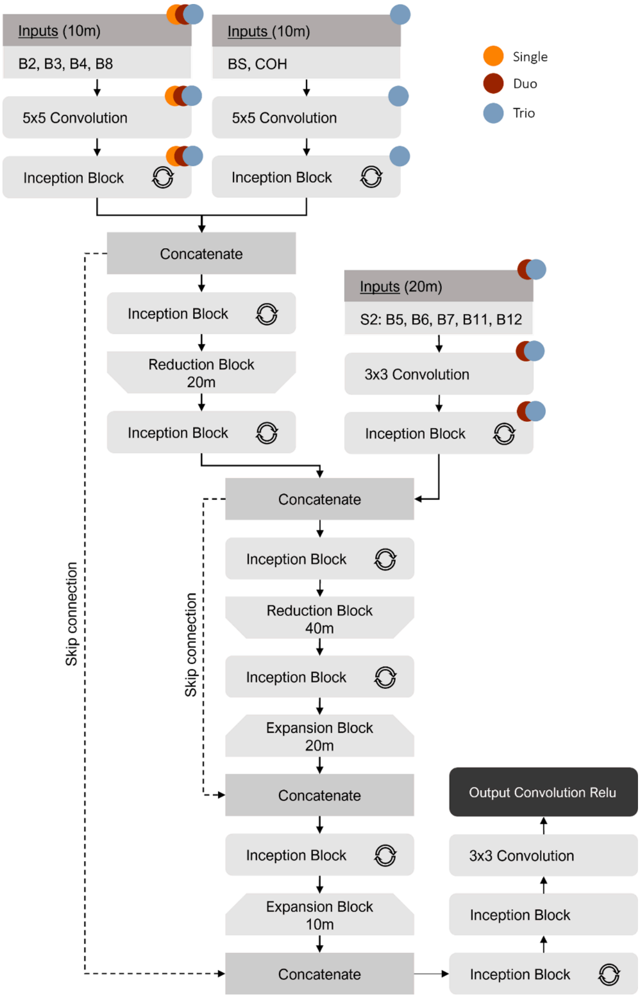


Fig. 11. The overall model architecture.

The model parameters were changed between the model input variable tests, the tile size tests, and the final large models. The large models have increased depth and width by an additional repetition of the inception layers, and the filter sizes have been increased by 20%. The number of parameters in each model is shown in Table 6.

The model design builds on Szegedy et al. (2017). by using inception and reductions blocks to structure the model. The design of the reduction blocks used is shown in Fig. 9. For the reduction blocks, adding an average pooling layer improved performance.

The inception blocks improve performance by making the networks "wider" rather than deeper. The added width means additional layers to each convolutional block with distinct tasks. The layers do 1D pooling, regular 3 × 3 convolutions and dimensionality reduction steps, as shown in Fig. 10.

Before joining the stems of the network, a 5 × 5 convolution and an inception block are introduced to capture additional image context. Three versions of the network were trained, building on the same base trunk. The colours in Fig. 11 indicate which of the models use the respective branches of the network.

The final weights of the model training on the area labels were used to initialise the weights of the volume trained model, and the volume model initialised the people trained model.

Joining the network branches at different places was investigated and joining before upscaling provided the best results. Combining the SAR data and the 10 m Sentinel 2 bands into one data input was investigated, but this yielded worse results than merging before downscaling, before upscaling or after upscaling.

4.6. Training and data accessibility

All the scripts necessary to replicate the study and the scripts to download the required data are available at https://github.com/casperfibaek/sentinel-structureson an AGPL License. The Buteo toolbox is required for some of the processing, especially the preprocessing steps. An interactive map of the predictions and labels are available in the repository. A GPU with at least 24 GB RAM is required to train the model – predictions are possible without a GPU.

The models were trained on a computer with the following specifications:

CPU	Intel i9 9980HK 8/16 cores
RAM	64 GB

(continued on next page)

Table 7
Abbreviation key for the used data sources.

Shorthand	Full name
RGB	Sentinel 2: B2, B3, B4
RGBN	Sentinel 2: B2, B3, B4, B8
RE	Sentinel 2: B5, B6, B7
SWIR	Sentinel 2: B11, B12
S2	Sentinel 2: B2, B3, B4, B5, B6, B7, B11, B12
VVa	Sentinel 1: Backscatter VV Polarisation - Ascending orbit
VVd	Sentinel 1: Backscatter VV Polarisation - Descending orbit
VHd	Sentinel 1: Backscatter VH Polarisation - Descending orbit
VHa	Sentinel 1: Backscatter VH Polarisation - Ascending orbit
COH	Sentinel 1: Interferometric Coherence - Ascending and descending orbit

(*continued*)

OS	Windows 11 Beta and Ubuntu 20.04 LTS
GPU	Nvidia GeForce RTX 3090–24 GB Ram (External)

Training the three initial models took 1 h 22 m each on average, repeated for each tested combination of input variables. The trio models took the longest to train at 1 h 37 m on average. The large models took 17 h each on average to train.

5. Results

The results were derived in three steps: (1) First, we analysed the impact of different combinations of input data on the model's prediction performance. (2) Upon finding the optimal combination, the ideal tile size for the model was found. (3) Finally, four models were trained on all the available data, building on the preceding steps. Three of the models were trained on each of the target labels using all the SAR data, including coherence from both orbital directions, and one was trained only on the area target label using the backscatter in both polarisations from the ascending orbital direction. The last model was trained on data from all of Denmark, excluding test sites, over two years. Table 7 shows the abbreviations used in the tables showing the results.

5.1. Model inputs

Three models were trained for each dataset, corresponding to each of the target labels, and a tile size of 64 × 64 pixels was used to compare the input datasets. In Tables 8–10, the results have been normalised to the accuracy of using only the blue, green, and red bands of the Sentinel 2 satellite, as the initial focus is on the relative performance of the input datasets. The two best performing models are shown in a light grey highlight. Table 8 shows the errors of the models over the Holstebro test areas.

Sentinel 2 data alone performs better than Sentinel 1 data, but Sentinel 1 and 2 combined yield significantly improved results. For Sentinel 2, including the near-infrared (NIR) band is highly significant for the model's performance, which can be improved by incorporating the red-edge bands. Using both orbital directions have the most significant effect on accuracy for the Sentinel 1 combinations. Including coherence and VV and VH polarisation in the model inputs slightly increased the accuracy, but the best results came from using all available data.

The results for the urban area of Aarhus show similar patterns to that of the mixed Holstebro area. Increasing the number of included Sentinel

Table 8
Error comparison for the Holstebro mixed urban area. The model trained only on the RGB bands from Sentinel 2 is used as the baseline and is shown in bold with absolute values.

			Area		Volume		People	
Datasets	Model	Train. Time	MSE	MAE	MSE	MAE	MSE	MAE
RGB	**Single**	**01:28:08**	**52.193**	**1.665**	**1875.0**	**8.398**	**1.6006**	**0.0184**
RGBN	Single	01:14:06	85.7%	88.5%	87.6%	88.7%	94.1%	91.9%
RGBN + RE	Duo	01:31:43	83.6%	84.9%	91.3%	89.8%	87.9%	80.6%
RGBN + SWIR	Duo	01:31:04	81.6%	83.7%	85.5%	84.7%	96.4%	90.3%
S2	Duo	01:37:30	78.8%	84.0%	86.2%	85.3%	89.1%	83.3%
VVa	Single	01:16:08	258.7%	187.3%	220.9%	170.9%	148.0%	125.4%
VVa + VHa	Single	01:09:27	235.8%	181.4%	203.0%	168.7%	139.6%	114.8%
VVa + VVd	Single	01:09:18	190.7%	161.9%	165.2%	151.4%	131.1%	123.2%
VVa + Coha	Single	01:03:44	210.2%	180.4%	188.9%	157.1%	138.2%	132.2%
VVa + VVd + COH	Single	01:16:22	173.6%	154.6%	146.5%	138.8%	127.5%	121.8%
VVa + VHa + COHa	Single	01:08:52	222.9%	179.1%	194.0%	162.7%	136.9%	118.9%
S2 + VVa + VHa	Trio	01:34:43	66.9%	76.1%	71.2%	78.2%	90.7%	84.5%
S2 + VVa + VVd + COH	Trio	01:39:59	64.2%	77.0%	67.3%	75.2%	86.2%	85.8%

Table 9
Error comparison for the Aarhus dense urban area. The model trained only on the RGB bands from Sentinel 2 is used as the baseline and is shown in bold with absolute values.

	Area		Volume		People	
Datasets	MSE	MAE	MSE	MAE	MSE	MAE
RGB	**135.6**	**4.331**	**14654.5**	**31.568**	**32.5450**	**0.1120**
RGBN	84.2%	89.4%	89.9%	90.2%	100.4%	97.1%
RGBN + RE	86.2%	88.5%	98.5%	94.6%	94.4%	91.0%
RGBN + SWIR	84.9%	86.1%	94.9%	89.5%	85.7%	92.8%
S2	81.4%	86.3%	94.0%	90.1%	87.4%	90.4%
VVa	243.8%	198.1%	187.3%	172.7%	156.4%	135.3%
VVa + VHa	231.1%	196.6%	173.4%	174.9%	143.1%	129.6%
VVa + VVd	191.9%	162.3%	144.8%	144.5%	130.8%	124.2%
VVa + Coha	215.9%	185.6%	173.0%	157.2%	149.1%	133.6%
VVa + VVd + COH	177.1%	157.0%	138.3%	140.1%	142.9%	122.6%
VVa + VHa + COHa	228.5%	204.5%	168.8%	173.2%	144.5%	131.8%
S2 + VVa + VHa	68.4%	78.2%	71.0%	80.7%	87.4%	91.2%
S2 + VVa + VVd + COH	65.0%	77.8%	70.6%	78.4%	82.0%	89.5%

Table 10
Error comparison for the rural area of Samsø Island. The model trained only on the RGB bands from Sentinel 2 is used as the baseline and is shown in bold with absolute values.

	Area		Volume		People	
Datasets	MSE	MAE	MSE	MAE	MSE	MAE
RGB	**9.299**	**0.284**	**227.1**	**1.260**	**0.0541**	**0.0018**
RGBN	82.4%	83.3%	88.0%	86.9%	97.7%	78.9%
RGBN + RE	83.2%	83.3%	83.9%	86.9%	92.1%	70.8%
RGBN + SWIR	82.6%	82.5%	82.9%	83.5%	105.4%	80.4%
S2	77.8%	82.1%	78.3%	81.1%	88.5%	73.1%
VVa	148.8%	178.7%	156.0%	174.8%	98.0%	129.5%
VVa + VHa	145.8%	176.3%	153.1%	183.9%	90.9%	110.1%
VVa + VVd	145.8%	176.3%	153.1%	183.9%	90.9%	110.1%
VVa + Coha	137.4%	173.5%	141.3%	147.8%	99.9%	134.2%
VVa + VVd + COH	116.3%	129.8%	119.9%	123.3%	96.8%	108.2%
VVa + VHa + COHa	146.4%	195.1%	147.3%	181.2%	100.6%	130.1%
S2 + VVa + VHa	69.2%	76.0%	71.0%	78.2%	96.7%	68.7%
S2 + VVa + VVd + COH	62.6%	74.6%	63.2%	71.9%	95.6%	73.7%

Table 11
Overlaps and the offsets used.

Tile Size 10 m	Tile size 20 m	Offsets for 10 m inputs	Offsets for 20 m inputs	Tiles tested
128 × 128	64 × 64	(0, 32) (0, 64), (0, 96) (32, 32) (64, 64), (96, 6) (32, 0), (64, 0), (96, 0)	(0, 16) (0, 32), (0, 48) (16, 16) (32, 32), (48, 48) (0, 32), (32, 0), (0, 48)	27,201
64 × 64	32 × 32	(0, 16) (0, 32), (0, 48) (16, 16) (32, 32), (48, 48) (0, 32), (32, 0), (0, 48)	(0, 8) (0, 16), (0, 24) (8, 8) (16, 16), (24, 24) (0, 8), (16, 0), (0, 24)	107,068
32 × 32	16 × 16	(0, 8) (0, 16), (0, 24) (8, 8) (16, 16), (24, 24) (0, 8), (16, 0), (0, 24)	(0, 4) (0, 8), (0, 12) (4, 4) (8, 8), (12, 12) (0, 4), (12, 0), (0, 12)	424,024
16 × 16	8 × 8	(0, 4) (0, 8), (0, 12) (4, 4) (8, 8), (12, 12) (0, 4), (12, 0), (0, 12)	(0, 2) (0, 4), (0, 6) (2, 2) (4, 4), (6, 6) (0, 2), (4, 0), (0, 6)	1,687,648

Table 12
Comparing impact on errors of different tiles sizes and prediction overlaps. The 128 × 128 pixel tile size is used as the baseline and is shown in bold with absolute values. “−9” designates that 9 overlaps have been used

Target	Tile Size	Batch sizes	MSE-0	MAE-0	MSE-9	MAE-9
Holstebro	**128 × 128**	**16 & 8**	**28.208**	**1.122**	**27.648**	**1.105**
”	64 × 64	64 & 32	100.0%	100.4%	98.4%	99.4%
“	32 × 32	256 & 128	100.7%	100.2%	96.6%	97.5%
“	16 × 16	1024 & 512	106.2%	102.7%	96.6%	97.3%
Aarhus	**128 × 128**	**16 & 8**	**71.541**	**2.911**	**70.038**	**2.872**
“	64 × 64	64 & 32	104.7%	100.7%	104.2%	100.0%
“	32 × 32	256 & 128	108.4%	102.5%	102.8%	99.7%
“	16 × 16	1024 & 512	112.8%	103.4%	102.6%	98.0%
Samsoe	128 × 128	**16 & 8**	**6.225**	**0.213**	**6.147**	**0.211**
“	64 × 64	64 & 32	98.9%	101.2%	97.7%	100.5%
“	32 × 32	256 & 128	96.9%	99.0%	93.0%	96.4%
“	16 × 16	1024 & 512	98.2%	98.4%	91.9%	93.9%

2 bands provided less improvement than it did for the mixed area of Holstebro, and the SAR dataset's performance slightly increased against the spectral data. The SWIR trained model outperformed the red-edge trained models; combining the two with the ten-meter bands yielded the best Sentinel 2 results.

The pattern of decreasing importance of Sentinel 2 bands continues for the rural areas on the island of Samsoe. Sentinel 1 performs almost as well as the Sentinel 2 trained models when predicting population. Sentinel 2 band combinations continue to yield the best results, and combinations of Sentinel 1 and 2 continue to provide the best performing models.

5.2. Tile size

The best performing model was the S2 + VVa + VVd + COH model, which was then used to compare the effect of using different tile sizes. When the batch sizes were sized to ensure an equal number of pixels per batch, processing time remained close to equal at around two and a half hours for each of the four tile sizes.

The comparisons are raw predictions for the test areas and the same model's predictions with nine additional overlaps. The median prediction of the nine overlaps and the original prediction is used in the comparison. Comparing the raw predictions with the version that uses overlaps show if the tile sizes accumulate errors along boundary regions and if the predictions are stable. The overlaps tested are shown in Table 11, and Table 12 show the performance of the different tile sizes.

The models' performance does not seem to be highly affected by the chosen tile size. The smaller the tile size chosen, the more significant the performance increase of using prediction overlaps, which corresponds to the expectation that smaller tile sizes lead to significantly more border regions, which are more likely to contain poor predictions. In urban areas, smaller tile sizes without overlaps perform worse, and in rural areas, they perform better. Using 32 × 32 pixel tiles at 10 by 10 m resolution and nine prediction overlaps were chosen as the base for further predictions.

5.3. Large models

Four distinct models were trained using the two best combinations of input layers. The best combination of input data requires SLC data from the Sentinel 1 satellites, which was only processed for the Central Denmark Region, covering 13,008 km^2. The models requiring SLC data were trained with all three target labels and took 17 h on average to

Table 13
Accuracy summary for the larger models using SLC data.

	Area			Volume			People		
Test Area	TPE	MSE	MAE	TPE	MSE	MAE	TPE	MSE	MAE
Holstebro	−0.61%	27.60	1.108	−8.34%	1048.40	5.694	−12.46%	0.0108	0.013
Aarhus	−3.12%	72.70	2.918	−2.12%	7990.64	21.792	−25.40%	0.2116	0.082
Samsoe	−4.04%	5.13	0.195	−12.75%	128.40	0.886	12.99%	0.0004	0.001
Merged	−1.90%	22.45	0.903	−6.53%	1617.58	5.628	−21.76%	0.0367	0.018

Table 14
Binary accuracy summary for the larger models using SLC data.

Location	Target	Accuracy	Bal. Acc	Precision	Recall	F1
Aarhus	Area	0.9181	0.9353	0.6859	0.9614	0.8006
“	Volume	0.8992	0.9231	0.6441	0.9600	0.7710
“	Population	0.9121	0.9220	0.5924	0.9352	0.7254
Holstebro	Area	0.9702	0.9587	0.7096	0.9454	0.8107
“	Volume	0.9585	0.9543	0.6352	0.9493	0.7611
“	Population	0.9706	0.9418	0.5935	0.9105	0.7186
Samsoe	Area	0.9925	0.9440	0.5659	0.8946	0.6932
“	Volume	0.9893	0.9525	0.4797	0.9149	0.6294
“	Population	0.9942	0.9068	0.4662	0.8186	0.5941
Merged	Area	0.9736	0.9619	0.6826	0.9487	0.7940
“	Volume	0.9653	0.9588	0.6216	0.9515	0.7519
“	Population	0.9737	0.9464	0.5821	0.9170	0.7122

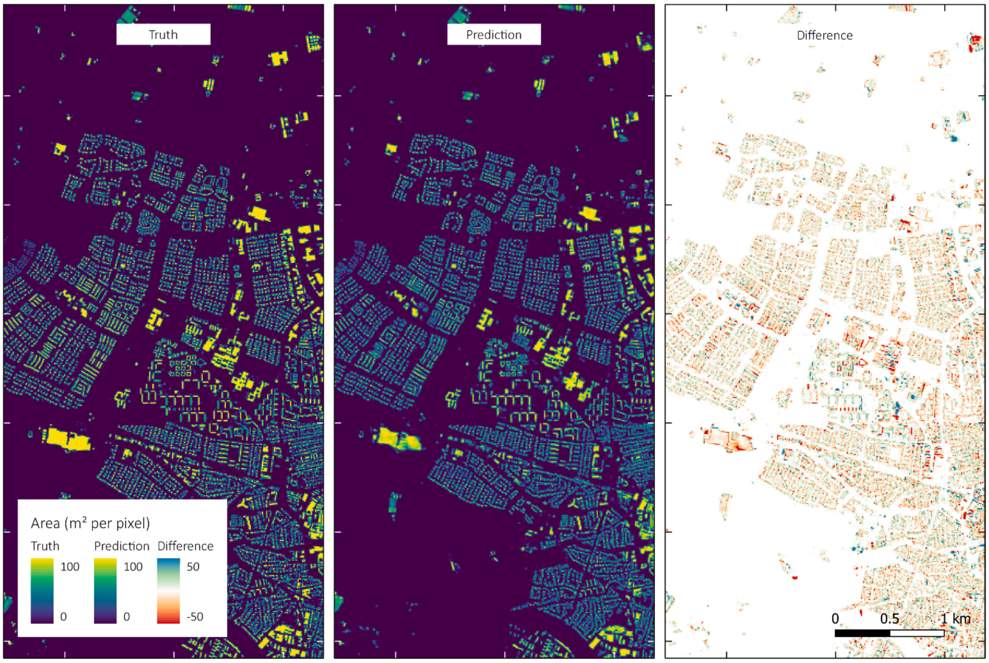


Fig. 12. Comparison between prediction and truth for building area.

train. The second-best version, which only requires GRD data from the Sentinel 1 satellites, was trained on data from two seasons across Denmark, excluding test sites, on the area label. The second model was trained on 1.7 million training samples and took 48 h to train. The model parameters were adjusted from the previous tests by adjusting the number of repeated inception blocks from two to three and increasing the filter count by 20%. The performance of the models trained with SLC data is shown in Table 13.

In Table 13, TPE is the total percentage error between the sum of the predicted values and the sum of the labels, divided by the sum of the labels. The merged test area is a virtual raster of the Holstebro, Aarhus and Samsoe test areas. The models consistently underestimate the total area, volume, and number of people but perform well at describing the spatial patterns of the structures. The increasing complexity of the model targets results in lower prediction accuracy. Converting the predictions to a binary classification on the existence of structural area, volume, or people on a given tile shows the model's high accuracy in replicating the general spatial distribution of structures. The binary accuracy is shown in Table 14, where area and volume predictions values above or equal to one are counted as positive. For the people labels and predictions, 0.001 people were used as the threshold. “Bal. Acc” refers to the class balanced accuracy metric as defined in Brodersen et al. (2010).

Figs. 12–14 show the predictions, labels and the difference between them. Fig. 12 shows the predictions produced by the area label trained model, shown for a zoomed-in area of the Holstebro mixed urban area,

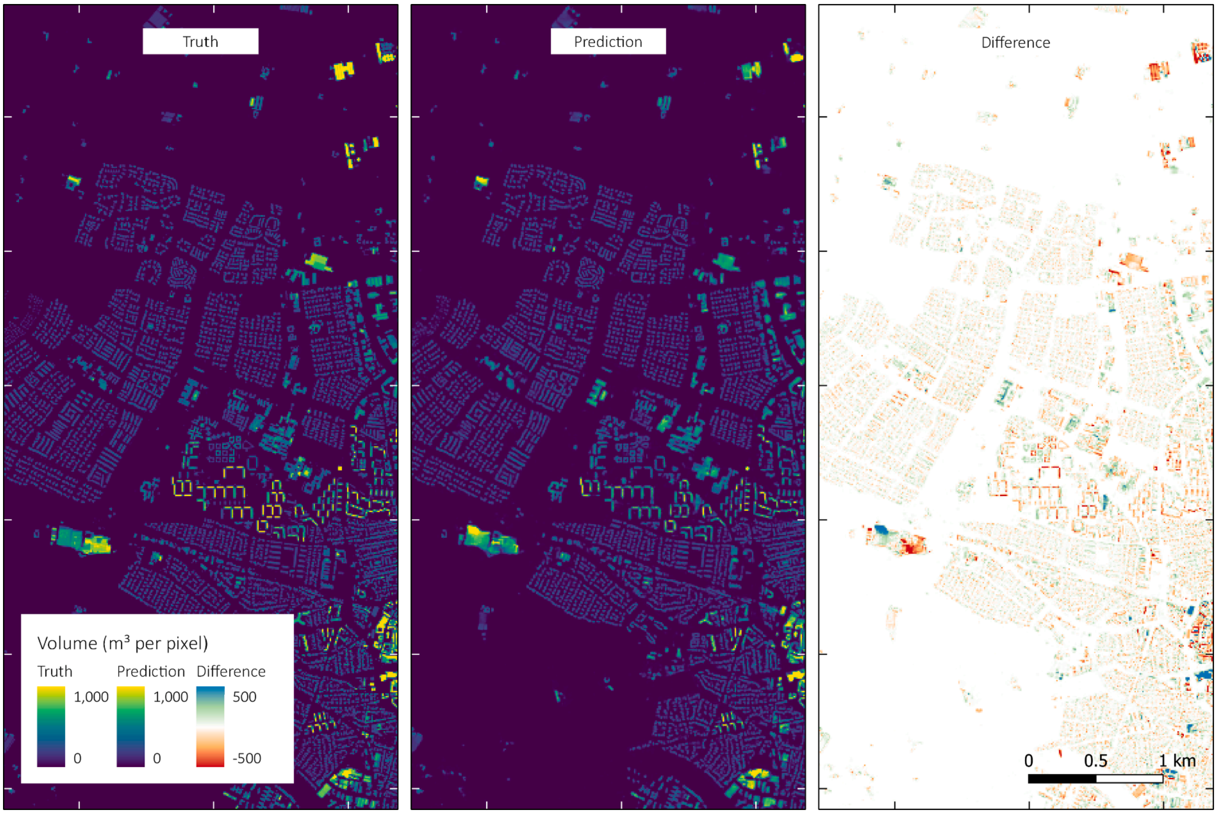


Fig. 13. Comparison between prediction and truth for building volume.

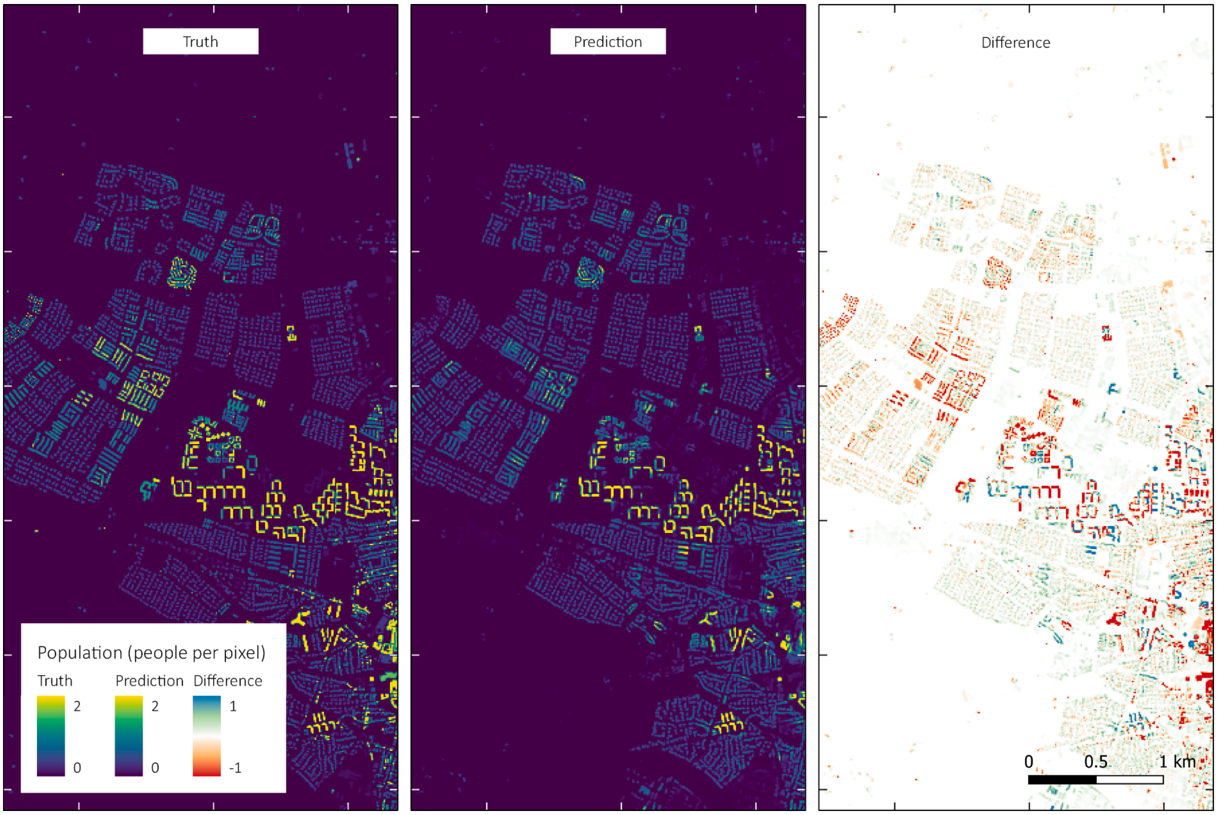


Fig. 14. Comparison between prediction and truth data for population.

Table 15
Results of the model trained on all of Denmark – without SLC data and Sentinel 1 GRD data from the descending orbital direction.

Test Area	Type	TPE	MSE	MAE	Accuracy	Bal. Acc	Precision	Recall	F1
Odense	Urban	−7.09%	66.487	2.281	0.9585	0.9150	0.8670	0.8529	0.8599
Bornholm	Rural	−14.92%	12.459	0.361	0.9931	0.8879	0.7886	0.7792	0.7839
Merged	Mixed	−9.41%	22.473	0.717	0.9867	0.9114	0.8418	0.8293	0.8355

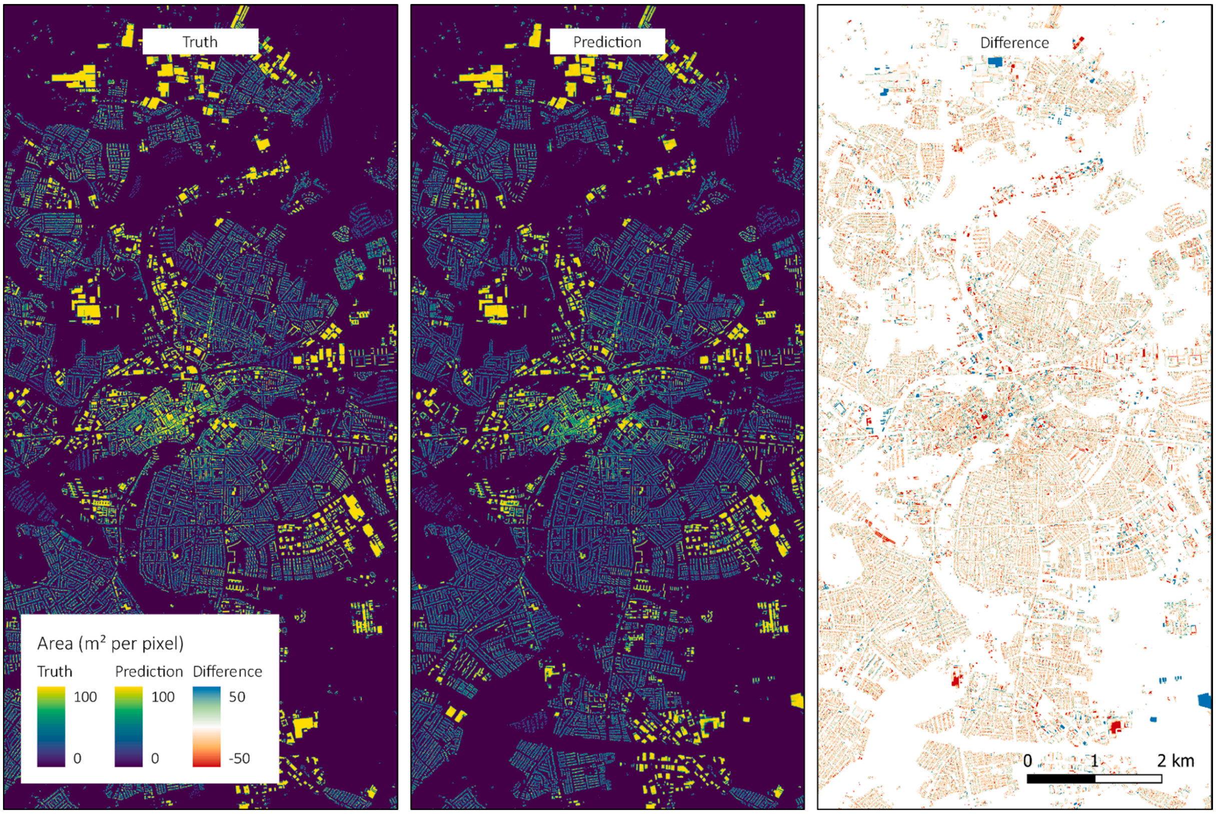


Fig. 15. Odense urban centre comparison.

showing the model's ability to reconstruct the general patterns of structures. The most significant errors in the scene are in the top right corner, where a scrapyard is confused with structures.

In Fig. 13, the difference in the predicted volume shows a very similar pattern to the errors in Fig. 12. The model has increasing issues with predicting the volume of the buildings in the urban core in the bottom right corner and issues with overpredicting one side of the warehouse in the bottom left and under predicting the other. As the volume of the building grows, so does the absolute error. The figures are displayed using the absolute errors and the absolute differences between the labels and predictions. This choice was made because displaying the errors relative to the maximum label value produced significant visual noise that made interpretation difficult, especially in suburban neighbourhoods due to pixels with small values.

The population predictions are shown in Fig. 14, which shows the most prominent errors, and interpreting the results is more complicated. The model does very well at reducing the influence of industrial and farm buildings but also removes some of the horseshoe-shaped residential blocks in the centre of the Figure.

Finally, a model was trained without data from the descending orbital direction and coherence from the Sentinel 1 constellation. The needed input variables are available globally, and less processing power is required due to the removal of the need to process SLC data. The model was trained on data from all Denmark, excluding test areas, for summer 2020 and spring 2021 using the area label target. The test areas for this model were Odense and Bornholm in summer 2020. Training the model took 48 h, and Table 15 and Figs. 15–17 show the predictions. The merged test area in Table 15 refers to a virtual raster combining the Odense and Bornholm test sites.

The model does well at describing the general patterns of the labels; a close-up is shown in Fig. 17. The most significant errors in Fig. 15 are the two industrial buildings in the bottom right and the greenhouses in the top. The binary accuracy is very high, while the model still consistently underestimates the total structural area. The underprediction is increasingly pronounced the more rural a target area is, as shown in Fig. 16.

Some smaller buildings in the rural areas are not predicted, which is likely due to vegetation overshadowing parts of the structures. As there are fewer structures in rural areas, missing a structure will increase the TPE considerably if the test area only covers rural areas.

6. Discussion

Our presented methodology on mapping structural characteristics using Sentinel 1 and 2 data is successful in replicating structural characteristics at a high level of accuracy. While the accuracy of the predictions decreases with the increased complexity of the target characteristics, the approach enables future predictions at the temporal resolution offered by the Sentinel satellites and a 10 by 10 m spatial resolution. Generating time series of structure characteristics will make

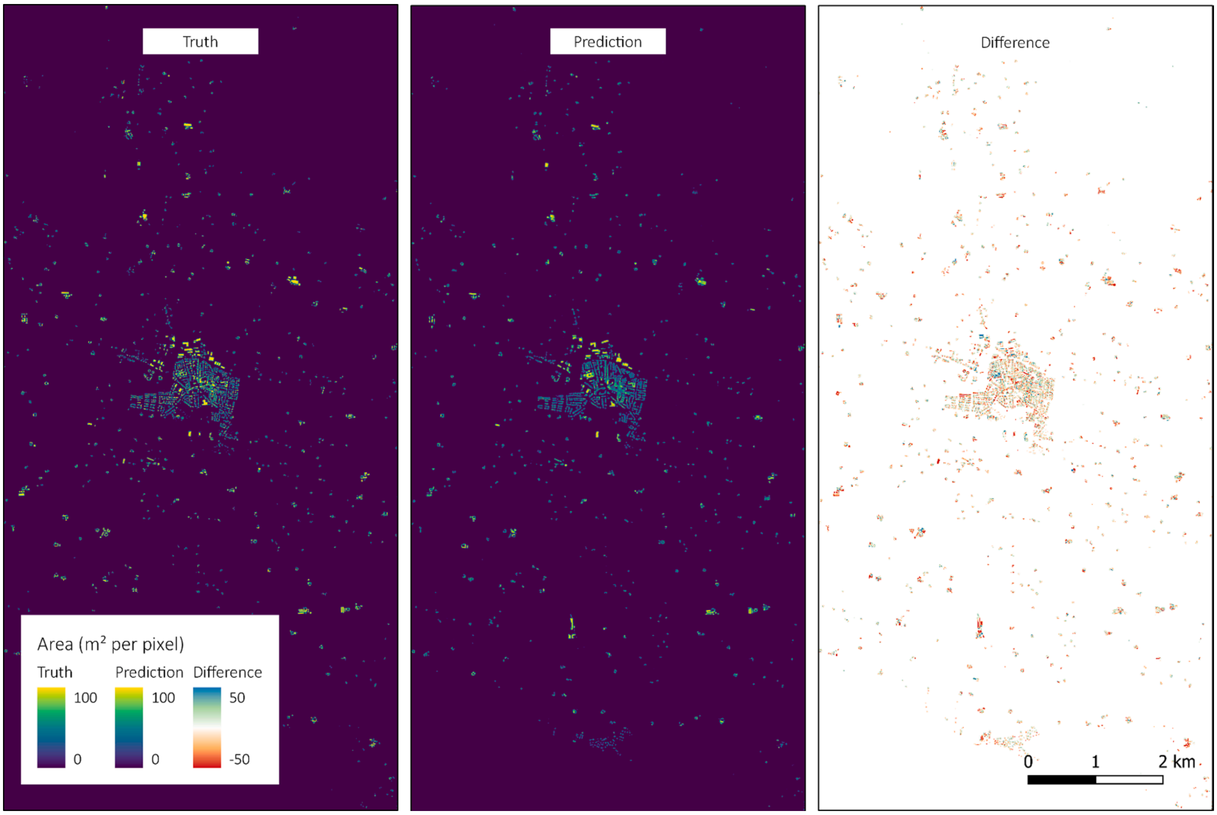


Fig. 16. Bornholm rural village comparison.

it possible to train improved deep learning models or create rules for cellular automata models to predict the spatial extent of urbanisation and guide urban planning as outlined by (Arsanjani et al., 2018).

The predicted structural characteristics have an MAE of less than 2.3 m^2 for area predictions in urban areas and an MAE of between 5.7 m^3 and 21.8 m^3 for mixed and urban areas volume predictions. Population predictions were made with an MAE of 0.08 people per pixel. Predictions on inhabitants of structures showed the highest total percentage error at −25.4% in dense urban areas and −12.5% in mixed areas.

Sentinel 2 imagery is the most critical data source for the models to achieve good performance, but relying on Sentinel 2 alone, means a dependency on cloud-free mosaics, which is an issue in time-sensitive mapping and areas that experience perpetual cloud cover. Interferometric coherence from Sentinel 1 increases the accuracy of the model. However, considering the computational efforts required to process and analyse coherence, leaving it out of the model, depending on the use case, is recommended. Combining ascending and descending imagery significantly boosts the models' performance, especially if Sentinel 1 is the only data source. The improved accuracy from including both orbital directions was also shown in Frantz et al. (2021), and this study's findings confirm their results. In cases where Sentinel 1 is combined with Sentinel 2, the improvements in accuracy by using both orbital directions is less pronounced. Combining Sentinel 1 and 2 data yielded the best results in all tests.

The total percentage errors for all tests were negative, which means consistent under predictions. A solution to the underpredictions could be adding an additional postprocessing network to the predictions to reduce the total percentage error. However, such an approach would not increase the number of structures captured by the model but scale the mean of the per tile predictions so that the overall sum of structural area, volume or population is more accurate at the expense of per-pixel accuracy. Accuracy might be improved by employing noisy student semi-supervised training iterations as done in Sirko et al. (2021) and described in Xie et al. (2020), which showed promising results. Adding textures to the input layers was not tested and might increase the accuracy of the models by adding additional context to the pixels, especially in border regions. However, as nine prediction overlaps were used for each prediction, the added benefit to border accuracy by textures is expected to be small. There was no dimensionality reduction done during the preprocessing of the input variables, such as a principal components analysis, which might have increased the models' performance. However, in the Inception-Resnet style architecture used in the study, the 1 × 1 convolutions in the inception blocks are used to perform dimensionality reduction (Szegedy et al. 2017).

The proposed models' geographical applicability could be extended by incorporating data from other geographical areas where high-quality terrain models and building footprints are available. A good candidate for collecting other ground truth data is the United Kingdom, where building footprints and terrain models are freely available. OpenStreetMap data is a good source of data for global footprints and has been proven to be a good auxiliary data source for population estimates (Stevens et al. 2015). Areas that have recently been mapped for building footprints, such as Nepal, Haiti, Uganda and Tanzania, which were recently updated due to efforts by the Humanitarian OpenStreetMap Team and contributors, would improve model performance. While Denmark is geographically homogenous, preliminary testing shows that the models presented in this study significantly boost the accuracy of models applied in Ghana when Deep Transfer Learning is methodologies are applied.

As the data sources used to generate the population ground truth data are unavailable in many countries, other approaches to generating population predictions from sentinel imagery should be investigated. Combining the model trained on the area or volume labels with a classification of structure– or settlement types and census information on housing patterns could be used to generate the population estimates. An approach like this is described by Leasure et al. (2020) for Ghana using WorldPop and Microsoft (2021) data. Applying their methodology for population estimations to the models proposed in this study could make

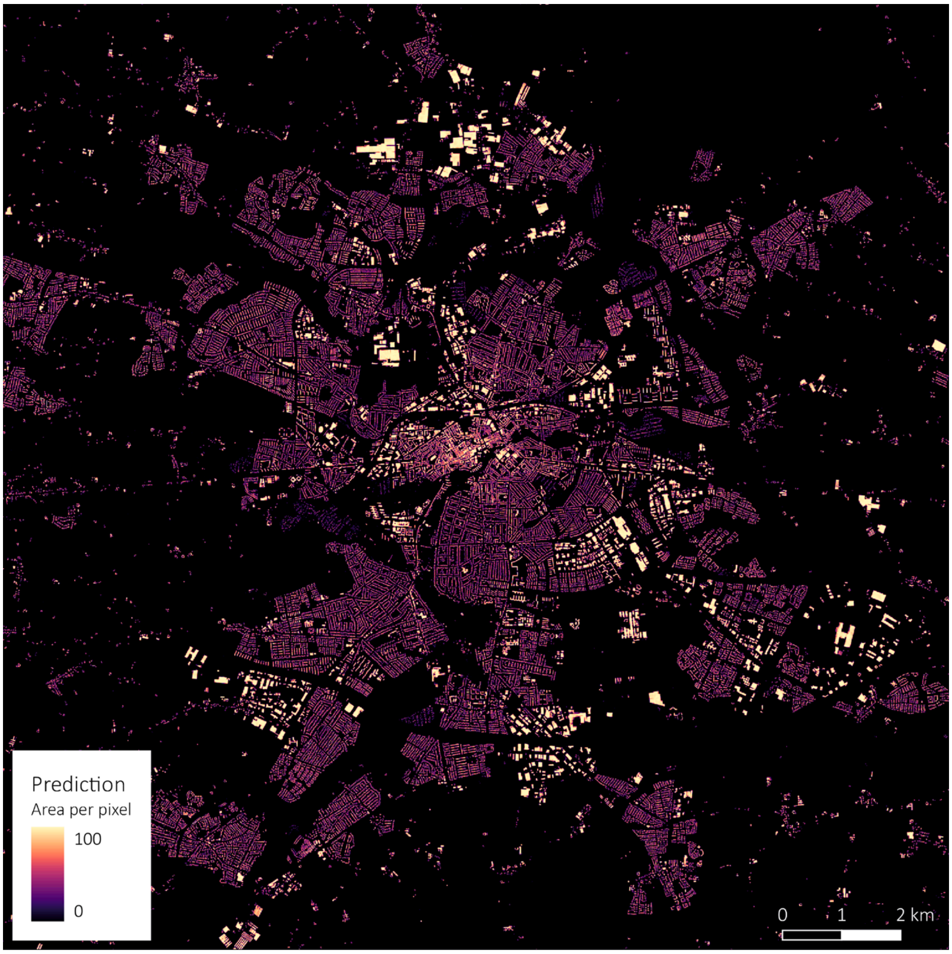


Fig. 17. Close up of the prediction of the structural area for the test area of Odense made without Sentinel 1 SLC data and GRD data from the descending orbital direction.

the population estimates from Sentinel 1 and 2 scalable to the global level.

7. Conclusion

The objective of this study was to investigate the possibility of mapping structures and their characteristics at 10 m spatial resolution using publicly available satellite data from the Copernicus Programme. We investigated different input variables' and tile sizes' effects on prediction performance using a multi-sensor approach with an Inception-ResNet style neural network to make predictions. The models perform well at describing the general spatial patterns of human-made structures and predicting the area and volume of structures. Population predictions show good promise but consistently underestimate the total population. The underprediction can be alleviated by scaling the predictions using volume or area predictions to lower the total percentage error or adding a postprocessing neural network. The ability to extract building features from the Sentinel satellites can likely improve the classification of urban settlement types by tying the classes to structure density and type. The large model, trained on area labels from all of Denmark during multiple seasons, was designed to have globally available input variables. We suggest further research is conducted on applying the models in more diverse geographical locations.

Funding

Casper is funded by the Danish Innovation Fund and the Niras/ Alectia Foundation.

CRediT authorship contribution statement

Casper Samsø Fibæk: Conceptualization, Methodology, Investigation. **Carsten Keßler:** Supervision. **Jamal Jokar Arsanjani:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

None.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2021.102628.

References

Abdikan, S., Sanli, F.B., Ustuner, M., Calò, F., 2016. Land cover mapping using sentinel-1 SAR data. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. International Society for Photogrammetry and Remote Sensing, pp. 757–61.

Agency European Space, ESA, 2021. Science Toolbox Exploitation Platform. SNAP Download. http://step.esa.int/main/ (September 9, 2021).

Jokar Arsanjani, J., Fibæk, C.S., Vaz, E., 2018. Development of a cellular automata model using open source technologies for monitoring urbanisation in the global south: the case of Maputo, Mozambique. Habitat Int. 71, 38–48.

Bowers, Samuel, 2021. Sen2mosaic — Bitbucket. https://bitbucket.org/sambowers/sen2mosaic/src/master/sen2mosaic/ (September 9, 2021).

Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M., 2010. The balanced accuracy and its posterior distribution. In: Proceedings - International Conference on Pattern Recognition, pp. 3121–3124.

Danish Meteorological Institute, 2019. Danish Climate. https://www.dmi.dk/fileadmin/user_upload/Rapporter/TR/2020/DMIRap20-01.pdf (September 9, 2021).

Ecopia, 2021. Digitize Africa. https://www.ecopiatech.com/africa (September 9, 2021).

Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., Strano, E., 2017. Breaking new ground in mapping human settlements from space – the global urban footprint. ISPRS J. Photogramm. Remote Sens. 134, 30–42.

European Space Agency, 2015. SENTINEL-1 Observation Scenario. European Space Agency. https://sentinel.esa.int/web/Sentinel/missions/Sentinel-1/observation-scenario (September 9, 2021).

Samsø Fibæk, C., Laufer, H., Keßler, C., Jokar Arsanjani, J., 2021. Geodata-driven approaches to financial inclusion – addressing the challenge of proximity. Int. J. Appl. Earth Observ. Geoinf. 99, 102325. https://doi.org/10.1016/j.jag.2021.102325.

Flatman, Andrew, et al., 2016. Quality Assessment Report to the Danish Elevation Model (DK-DEM) Quality Assessment Report to the Danish Elevation Model (DK-DEM) Agency for Data Supply and Efficiency. http://grunddata.dk (September 9, 2021).

Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., van der Linden, S., Hostert, P., 2021. National-scale mapping of building height using sentinel-1 and sentinel-2 time series. Remote Sens. Environ. 252, 112128. https://doi.org/10.1016/j.rse.2020.112128.

Gao, Yan, Cui, Yan, 2020. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. Nat. Commun. 11 (1), 1–8 https://www.nature.com/articles/s41467-020-18918-3 (September 9, 2021).

GeoDenmark, 2021. GeoDanmark Specification 6.0.1. http://www.geodanmark.nu/Spec6/HTML5/DK/601/StartHer.htm (September 9, 2021).

Herfort, Benjamin, et al., 2021. The evolution of humanitarian mapping within the openstreetmap community. Sci. Rep. 11 (1), 1–15 https://www.nature.com/articles/s41598-021-82404-z (September 9, 2021).

Hyndman, Rob J., Athanasopoulos, George, 2018. Forecasting : Principles and Practice. 3rd ed. OTexts.

Jacob, A.W., Vicente-Guijalba, F., Lopez-Martinez, C., Lopez-Sanchez, J.M., Litzinger, M., Kristen, H., Mestre-Quereda, A., Ziolkowski, D., Lavalle, M., Notarnicola, C., Suresh, G., Antropov, O., Ge, S., Praks, J., Ban, Y., Pottier, E., Mallorqui Franquet, J. J., Duro, J., Engdahl, M.E., 2020. Sentinel-1 InSAR coherence for land cover mapping: a comparison of multiple feature-based classifiers. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 535–552.

Jasper, Paul, Hernandez, Andrea Luna, Scott, Molly, Tzavidis, Nikos, 2021. How can new technology support better measurement of extreme poverty? Annex A - methods for producing high-resolution and high-frequency poverty estimates. DEEP.

Jiang, L., O'Neill, B.C., 2017. Global urbanization projections for the shared socioeconomic pathways. Global Environ. Change 42, 193–199.

Khabbazan, Saeed, et al., 2019. Crop monitoring using sentinel-1 data: a case study from the Netherlands. Remote Sensing 11 (16), 1887 https://www.mdpi.com/2072-4292/11/16/1887/htm (September 9, 2021).

Kingma, Diederik P., Ba, Jimmy Lei, 2015. Adam: a method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, International Conference on Learning Representations, ICLR. https://arxiv.org/abs/1412.6980v9 (September 9, 2021).

Koppel, K. et al., 2015. Sentinel-1 for urban area monitoring - analysing local-area statistics and interferometric coherence methods for buildings' detection. In: International Geoscience and Remote Sensing Symposium (IGARSS). Institute of Electrical and Electronics Engineers Inc., pp. 1175–78.

Leasure, D.R., Jochem, W.C., Weber, E.M., Seaman, V., Tatem, A.J., 2020. National population mapping from sparse survey data: a hierarchical bayesian modeling framework to account for uncertainty. PNAS 117 (39), 24173–24179.

Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49 (4), 764–766.

Li, Weijia, et al., 2019. Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. Remote Sensing 11 (4), 403 https://www.mdpi.com/2072-4292/11/4/403/htm (September 9, 2021).

Li, Wenwen, Goodchild, Michael F., Church, Richard, 2013. An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. Int. J. Geogr. Inf. Sci. 27 (6), 1227–1250 https://www.tandfonline.com/doi/abs/10.1080/13658816.2012.752093 (September 9, 2021).

Li, X., Du, Z., Huang, Y., Tan, Z., 2021. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. ISPRS J. Photogramm. Remote Sens. 179, 14–34.

Li, X., Zhou, Y., Gong, P., Seto, K.C., Clinton, N., 2020. Developing a method to estimate building height from sentinel-1 data. Remote Sens. Environ. 240, 111705. https://doi.org/10.1016/j.rse.2020.111705.

Microsoft, 2021. Open Dataset of Machine Extracted Buildings in Uganda and Tanzania. https://github.com/microsoft/Uganda-Tanzania-Building-Footprints (September 9, 2021).

Mueller-Wilm, Uwe, 2021. Sen2Three Processor. https://github.com/senbox-org/sen2three (September 9, 2021).

Panek, Jiri, Netek, Rostislav, 2019. Collaborative mapping and digital participation: a tool for local empowerment in developing countries. Information (Switzerland) 10 (8), 255 https://www.mdpi.com/2078-2489/10/8/255/htm (September 9, 2021).

Qi, Wei, Liu, Shenghe, Gao, Xiaolu, Zhao, Meifeng, 2015. Modeling the spatial distribution of urban population during the daytime and at night based on land use: a case study in Beijing, China. J. Geogr. Sci. 25 (6), 756–768 https://link.springer.com/article/10.1007/s11442-015-1200-0 (September 9, 2021).

Qiu, C., Schmitt, M., Geiß, C., Chen, T.-H., Zhu, X.X., 2020. A framework for large-scale mapping of human settlement extent from sentinel-2 images via fully convolutional neural networks. ISPRS J. Photogramm. Remote Sens. 163, 152–170.

Semenzato, Andrea, et al., 2020. Mapping and monitoring urban environment through sentinel-1 SAR data: a case study in the Veneto region (Italy). ISPRS Int. J. Geo-Inf. 9 (6), 375 https://www.mdpi.com/2220-9964/9/6/375/htm (September 9, 2021).

Sirko, Wojciech, et al., 2021. Continental-scale building detection from high resolution satellite imagery. arXiv. https://arxiv.org/abs/2107.12283v2 (October 16, 2021).

Smith, S.L., Kindermans, P.J., Ying, C., Quoc, V.L., 2018. Don't decay the learning rate, increase the batch size. 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, International Conference on Learning Representations, ICLR.

Stevens, Forrest R., Gaughan, Andrea E., Linard, Catherine, Tatem, Andrew J., 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PLoS One 10 (2), e0107042 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0107042 (September 9, 2021).

Szegedy, Christian, Ioffe, Sergey, Vanhoucke, Vincent, Alemi, Alexander A., 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: 31st AAAI Conference on Artificial Intelligence, AAAI 2017, pp. 4278–84. https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806 (September 9, 2021).

Tan, Mingxing, Le, Quoc V., 2019. EfficientNet: rethinking model scaling for convolutional neural networks. In: 36th International Conference on Machine Learning, ICML 2019, PMLR, 10691–700. https://proceedings.mlr.press/v97/tan19a.html (September 9, 2021).

Tatem, Andrew J., 2017. WorldPop, open data for spatial demography. Sci. Data 4 (1), 1–4 https://www.nature.com/articles/sdata20174 (September 9, 2021).

Trouvé, Emmanuel, Chambenoit, Yoann, Classeau, Nicolas, Bolon, Philippe, 2003. Statistical and operational performance assessment of multitemporal SAR image filtering. IEEE Trans. Geosci. Remote Sens. 41 (11), 2519–2530.

Woon, Chih Yuan, 2013. The Global South. In: The Ashgate Research Companion to Critical Geopolitics, SAGE PublicationsSage CA, Los Angeles, CA, pp. 323–40. https://journals.sagepub.com/doi/10.1177/1536504212436479 (October 18, 2021).

Xie, Q., Luong, M.T., Hovy, E., Quoc, V.L., 2020. Self-training with noisy student improves imagenet classification. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 10684–10695.

Xu, Bing, Wang, Naiyan, Chen, Tianqi, Li, Mu, 2015. Empirical evaluation of rectified activations in convolutional network. arXiv. https://arxiv.org/abs/1505.00853v2 (September 9, 2021).

Zhang, Michael R., Lucas, James, Hinton, Geoffrey, Ba, Jimmy, 2019. Lookahead Optimizer: K Steps Forward, 1 Step Back. In: Advances in Neural Information Processing Systems, Neural information processing systems foundation. https://arxiv.org/abs/1907.08610v2 (September 9, 2021).

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Cham, pp. 3–11.