A deep learning method for creating globally applicable population estimates from sentinel data

Fibæk, Casper Samsø; Keßler, Carsten; Arsanjani, Jamal Jokar; Trillo, Marcia Luz

Published in:
Transactions in GIS

DOI (link to publication from Publisher):
10.1111/tgis.12971

Publication date:
2022

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

DOI: 10.1111/tgis.12971

RESEARCH ARTICLE

Transactions in GIS WILEY

A deep learning method for creating globally applicable population estimates from sentinel data

Casper Samsø Fibæk[1,2] | Carsten Keßler[1,3] | Jamal Jokar Arsanjani[1] | Marcia Luz Trillo[1,2]

[1]Department of Planning, Aalborg University, Copenhagen, Denmark

[2]NIRAS Consulting, Allerød, Denmark

[3]Department of Geodesy, Bochum University of Applied Sciences, Bochum, Germany

Correspondence
Casper Samsø Fibæk, Department of Planning, Aalborg University, Copenhagen, A.C. Meyers Vænge 15, Copenhagen 2450, SV, Denmark.
Email: cfi@plan.aau.dk

Funding information
Innovationsfonden; NIRAS Alectia Foundation

Abstract

Recent research has shown promising results for estimating structural area, volume, and population from Sentinel 1 and 2 data at a 10 by 10-m spatial resolution. These studies were, however, conducted in homogeneous countries in Northern Europe. This study presents a deep learning methodology for population estimation in areas geographically distinct from Northern Europe. The two case study areas are Ghana and Egypt's Mediterranean coast, with supplementary ground truth data collected from Uganda, Kenya, Tanzania, Palestine, and Israel. This study aims to answer the question: How can we use Deep Learning to map structural area and type to derive population estimates for Ghana and Egypt based on Sentinel data? At 10 by 10-m resolution, the accuracy of the presented area predictions is similar to the Google Open Buildings dataset. An intercomparison of the presented population predictions is made with global state-of-the-art spatial population estimates, and the results are promising, with the proposed methodology showing comparable or better results than the state-of-the-art for the study areas.

1 | INTRODUCTION

People are increasingly living in cities, and urbanization is taking place rapidly. By 2050, the UN projects that 68% of the world's population will live in urban areas compared with 30% in 1950 (United Nations, 2019). While cities only cover 0.5% of the world's land, this has doubled since 1975. In Northern Africa, land occupied by cities has quadrupled since 1975, and in sub-Saharan Africa, it has tripled (Dijkstra et al., 2021). Urban areas are the causes of a significant amount of global carbon emissions (Moran et al., 2018), but while living in dense cities, generally reduces the carbon footprint of the individual, different urban structures impact the environment in distinct ways (Høyer & Holden, 2003; Li et al., 2013). Global maps of human settlements with population estimates enable spatial modeling of household emissions, consumption, and service levels (Jiang & O'Neill, 2017).

This study is a continuation of the work described in Fibæk, Keßler, et al. (2021), which presents a multisensor approach that can reach high levels of accuracy for mapping the area, volume, and population of structures in Denmark. That research, in turn, builds on Frantz et al. (2021) and Haberl et al. (2021), which showed that Sentinel imagery could be used to predict the height of structures, as well as Corbane et al. (2020) and the Global Human Settlement Layers' S2-NET, which has been used to create global maps of built-up areas extracted using the Sentinel 2 satellites and Deep Learning. The European Space Agency and Esri have also published global land cover classifications for 2020 based on Sentinel data, which features a built-up class (ESA, 2020; Esri, 2020). These datasets feature binary classifications of built-up areas and do not predict the physical characteristics of the present structures nor estimate the population.

Training deep learning models in areas where significant ground truth data are available, such as Denmark, and adapting them to different local contexts could reduce the information inequalities and improve decision-making (Fibæk, Laufer, et al., 2021). Sentinel 1 and 2 data is well suited for far-field applications as global data are openly and freely available. The spatial resolution of the Sentinel satellites' data products allow the mapping of individual structures and characteristics, such as area, volume (Haberl et al., 2021) and estimated population (Fibæk, Keßler, et al., 2021; Schug et al., 2021), where structures refer to buildings and other human-made space-enclosing structures. However, current approaches to structure mapping from Sentinel data have either not shown applicability in data-poor regions (Fibæk, Keßler, et al., 2021; Schug et al., 2021) or do not utilize the full potential of multisensor approaches nor all the spectral bands available from the Sentinel 2 satellite (Corbane et al., 2020).

Multiple studies have been conducted on extracting building footprints from very high-resolution satellite imagery using deep learning models (Weijia Li et al., 2019) and made available through humanitarian programs—such as “Google Open Buildings” (Sirko et al., 2021), Meta (Facebook) “Data for Good” population estimates (Meta, 2022) and Microsoft “AI for Good” (Microsoft, 2021). Using the proposed models and gathering the required very high-resolution satellite imagery at the country or global scale to extract structural information can be infeasible due to processing requirements, low temporal resolution, or financial costs unless supported by donations or not continuously updated. Furthermore, the results section shows that they currently do not detect a significant number of structures and population clusters in the study areas, especially in self-organized areas.

WorldPop has created global population maps at different scales using a combination of approaches, including one based on building footprints from Ecopia and Maxar Technologies made available at 100 by 100 m resolution (Ecopia, 2021; Stevens et al., 2015). Meta provides 30 by 30-m resolution predictions based on counting the number of buildings per pixel derived from very high-resolution imagery (Meta, 2022). The WorldPop data provide many different aggregates of the human population, such as constraining the predictions to UN population forecasts. For parts of the study areas, the data used to generate the cloud-free mosaics of very high-resolution satellite imagery underpinning the models is dated or missing, which can cause significant planning issues considering the pace of changes in urban settlements. Current population estimates from WorldPop and Meta do not differentiate between daytime and nighttime population (Foley, 1954) for the study areas, which is important for

mapping population dynamics. WorldPop has mapped population dynamics in France through cellphone tower activity, but not through Earth Observation alone (Deville et al., 2014), which is investigated in this study. The results section of this study compares the WorldPop, and Meta population estimates and the Sentinel derived population estimates presented in this study, nicknamed SenPop.
The main contribution of this article can be summarized as follows:

- Shows a viable multisensor approach for predicting structural area at a 10 by 10-m spatial resolution applied in Ghana and Egypt.
- Provides a methodology for converting the structural characteristics and census data into day- and night-time population maps.
- Shows the viability of training Earth observation deep learning models on ground truth data from Denmark to locations where less ground truth data is available, using minor amounts of additional local training data.

The remainder of this article is organized in the following way: Section 2 introduces the study areas and the datasets used to generate labels. Section 3 describes the preprocessing and data augmentation steps taken to ready the data for ingestion into the deep learning model. Section 4 describes the methodology behind the predictions: (1) Structural area per pixel, (2) Structure type, (3) Population per pixel, and (4) Structural volume per pixel. Section 5 presents the results and comparisons with other datasets on structural area and population estimates. Section 6 discusses the results, and Section 7 summarizes the article's conclusions.

2 | STUDY AREA

The two study areas are Ghana and Egypt's Mediterranean coast. Training and testing data were collected from Ghana, Denmark, Gaza, and parts of Egypt, Israel, Palestine, Tanzania, Kenya, and Uganda. Predictions on area, structure type, and population were made for the two primary study areas. The choice of including Danish ground truth data is due to the availability of complete ground truth data and the previous application of Sentinel data to map structural characteristics in Denmark (Fibæk, Keßler, et al., 2021; GeoDenmark, 2021). Including data from Northern, Western, and Eastern Africa ensures the applicability of the models in diverse contexts. All the data used as training data were sensed between Q2 and Q4 2021. Table 1 shows the origin of the training data collected on structural footprints. Figures showing the case study areas, and areas where supplementary training and test data were collected, are included in Appendix A.

2.1 | Ghana

Ghana is a West African country sharing borders with Burkina Faso to the North, Cote d'Ivoire to the West, Togo to the East and the Gulf of Guinea and the Atlantic Ocean to the South. The country has a population of 30.8 million people (Ghana Statistic Office, 2021) and covers 239,567 km^2. According to the updated Köppen-Geiger classification, Ghana consists of two climate zones, tropical monsoon in the southwest and tropical Savannah elsewhere (Peel et al., 2007). Having different climate zones in the study area means that any deep learning model trained to predict structural characteristics needs to be able to generalize across climate types. Going from the South to the North as it nears the Sahel Region, Ghana is increasingly arid. The main population centers are the Metropolitan areas of Accra and Kumasi. The coastal region has highly consistent cloud cover, especially the western parts of the coastline from Cape Coast toward the Cote d'Ivoire border. A new population and housing Census was conducted in 2021, which serves as the basis for the population distribution of the models. The census was initially scheduled for 2020 but postponed due to the outbreak of Covid-19,

TABLE 1 Study areas and training material

Country	Area	Purpose	Structures count (M)	Structures area (km^2)	Training area (km^2)	Data capture
Ghana	Distributed nationally	Training	0.40	46.77	4283	2021: Q3–Q4
"	Distributed nationally	Test	0.09	5.01	94	2021: Q3–Q4
"	Distributed nationally	Student	Varies	Varies	785	2021: Q3–Q4
Egypt	Mediterranean Coast	Training	0.05	8.80	469	2022: Q1
Egypt	Mediterranean Coast	Test	0.004	0.60	10	2022: Q1
Palestine	Gaza	Training	0.03	6.36	28	2022: Q1
Israel	Coast and the Negev	Training	0.05	21.90	323	2022: Q1
Tanzania, Kenya	Kilimanjaro Foothills	Training	0.77	46.29	10,552	2021: Q4
"	Mwanza[a]	Training	0.42	31.60	1322	2021: Q3–Q4
"	Kigoma, Tanganyika Coast	Training	0.09	3.26	3523	2021: Q4
"	Dar es Salaam & Zanzibar	Training	1.41	139.25	2613	2021: Q3–Q4
Uganda	Kampala, Entebbe & Jinja[a]	Training	1.46	120.09	3130	2021: Q3–Q4
Denmark	All[a]	Training	5.69	726.35	42,933	2021: Q1–Q4
Total (sum)			10.41	1156.28	69,596	

[a]Multiple timestamps used. (M) Refer to "millions".

2.2 | Mediterranean Egypt

Egypt's Mediterranean coast is home to the Nile Delta, the City of Alexandria, the Suez Canal, and diverse climate zones. The area is home to approximately 40 million people, and the Egyptian Government is currently undertaking projects on Integrated Coastal Zone Management as the area is susceptible to sea inundation, erosion, and saltwater intrusion (Abou-Mahmoud, 2021; Fabbri, 1998; NIRAS, 2021). A high-resolution map of structures and populations could provide a better outset for urban growth planning, economic and environmental impact assessments, and insurance valuations. Creating such maps using Sentinel data could be incorporated into continuous coastal zone monitoring efforts (Elnabwy et al., 2020).

Training data were collected along the entirety of Egypt's Mediterranean coast. Previous predictions of urban structures in Egypt, made using very high-resolution satellite Earth observation, have shown that it can be a challenging area to extract building footprints, likely due to the building materials used, the proximity of structures and walled courtyards (Sirko et al., 2021). The last census in Egypt was conducted in 2017, and the population estimates used in the study are based on UN forecasts from the 2017 census.

2.3 | Eastern Africa

In Eastern Africa, data were collected from Tanzania, the Zanzibar Archipelago, and the Tanzania border region of Kenya, and Uganda. The areas were chosen as they represent a mix of settlement types and diverse geographies and range from the mountainous Kilimanjaro foothills on the border of Tanzania and Kenya to the densely populated urban centers of Dar es Salaam, Tanzania and Kampala, Uganda. Mwanza, on the coast of Lake Victoria, also called the City of Rocks, was chosen due to its rocky landscape, which is often difficult to map using Earth Observations and Deep Learning (Sirko et al., 2021). The difficulty comes from the size, shadows cast, and the spectral and radar backscatter signature of rocky landscapes resembling housing, especially housing constructed using natural materials.

Tanzania and Uganda were initially chosen due to Microsoft releasing their building footprints dataset for those two countries (Microsoft, 2021). Then, areas within Tanzania and Uganda were selected that had good coverage of OpenStreetMap building footprints and provided diverse physical characteristics and geographies.

2.4 | Gaza, Palestine, and Israel

Supplementary training data were collected from Gaza, Palestine, and Israel. Gaza was chosen as a visual inspection showed a very high degree of coverage of OpenStreetMap (OSM) structures, which is potentially due to the efforts of Humanitarian OpenStreetMap (https://wiki.openstreetmap.org/wiki/Humanitarian_OSM_Team/Gaza). Like Gaza, Israel, and Palestine were selected due to Egypt's geographical proximity and the good OSM coverage along the coast and the hilly areas around Jerusalem.

2.5 | Denmark

Denmark was chosen due to the ease of accessing high-quality ground truth for the country for multiple seasons. Denmark is a largely geographically homogenous area, meaning that the far-field application areas of models for predicting structural characteristics trained on Danish data are likely to perform poorly without additional local training data.

3 | DATA COLLECTION AND PREPROCESSING

All input and ground truth data were processed using the Buteo, Orfeo, and ESA SNAP toolboxes through Python (Agency European Space, 2021; Fibaek, 2022; Grizonnet et al., 2017).

3.1 | Ground truth data

Ground truth data came from a mixture of predominantly OpenStreetMap and GeoDenmark data, supplemented with data from Google Open Buildings, using the 90% precision threshold (Sirko et al., 2021), and Microsoft Buildings for Tanzania and Uganda. All training sites were manually inspected, and missing structures were digitized using the latest available imagery from Google Earth and Microsoft Bing Maps. Ground truth data in Ghana was collected in collaboration with the Center for Remote Sensing and Geographic Information Services at the University of Ghana, Legon. Ground truth data for Egypt was gathered in cooperation with the University of Alexandria.

The vector training sites and buildings were rasterized to a 20 by 20 cm resolution, with 1.0 assigned to buildings and 0.0 otherwise. The resulting rasters were then resampled to 10 by 10 m cells using the average and multiplied by 100, resulting in the approximate m^2 coverage of structures within each pixel. Ground truth data on population came from the 2021 Ghana Population and Housing Census (Ghana Statistic Office, 2021) and the 2017 Egypt e-census and UN-Population projections (CAPMAS, 2017; United Nations Population Fund, 2021).

3.2 | Sentinel data

All available Sentinel 2 data with less than 20% cloud cover was downloaded for the data capture periods specified in Table 1 through ESA SciHub and the Onda DIAS. Sentinel 1 data were captured for 1 month in the middle of the Sentinel 2 data capture period. If it was not possible to generate a cloud-free mosaic with the downloaded imagery, the allowed cloud percentage was increased, and new data was downloaded iteratively until a cloud-free mosaic could be produced. The Sentinel 1 Ground Range Detected (GRD) data were processed through a standard GRD to Backscatter (dB) workflow in the Graph Processing Tool of ESA SNAP. The Sentinel 1 images were merged using the Buteo Toolbox, where multitemporal speckle filtering was done through the weighted median of an ellipsoidal kernel shown in Figure 1. Each pixel is weighted according to the overlap with an ellipsoid centered around the most

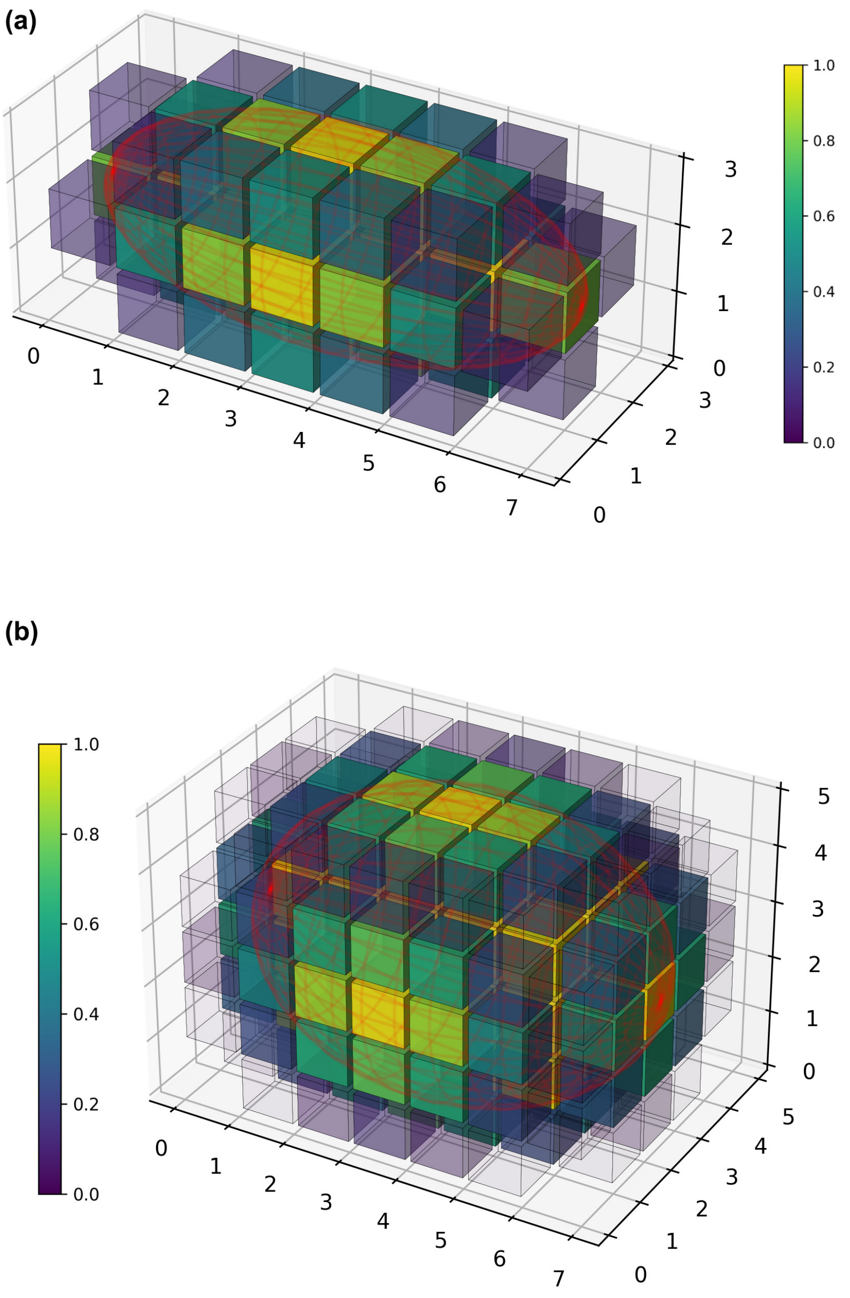

FIGURE 1 (a) Ellipsoidal 3×3×7 kernel. (b) Ellipsoidal 5×5×7 kernel. Figures show the kernels used in the weighted median preprocessing step. The shaded values (0–1) signify the attributed weights before normalization, and (0–7) represent imagery and timestamps, with the central timestamp located at 3 to 4.

temporally central image. The weighted median is then calculated from the values and weights using linear interpolation of the sorted and weighted values. The borders between the Sentinel 1 images were feathered using a 2 km distance. 32×32 pixel tiles were extracted from the resulting mosaics with 10 additional overlaps at different pixel offsets. The preprocessed VV and VH polarizations were normalized to [0, 1] after converting to dB and thresholding the images between −30 and 20 dB. The Sentinel 1 imagery was aligned with the Sentinel 2 images before tile creation.

The Sentinel 2 level 2A (atmospherically corrected, bottom-of-atmosphere reflectance) images were mosaiced using the Buteo toolbox, where a quality estimate image is made for each Sentinel 2 image, after which they are merged to ensure the highest overall quality with as few images used in the mosaic as possible. Images temporally further away than the image with the highest quality have their overall quality reduced with the temporal distance at the pace of a 1% quality reduction per week. The borders between images were feathered with a 1 km distance buffer. The images were normalized to [0, 1] after thresholding the images to between 0 and 10,000. 32×32 pixel tiles were extracted from the 10 m bands, and 16×16 pixel tiles were extracted for the 20 m bands covering the same areas as the 10 m bands.

The Sentinel products used are shown in Table 2:

4 | METHODOLOGY

The method used for extracting the various structural characteristics follows a modified approach to the work presented in Fibæk, Keßler, et al. (2021). This modified approach adds noisy student training steps for the model (Xie et al., 2020) and minor changes to the model design pre- and post-processing steps. The semi-supervised noisy student iterations were implemented, as shown in Figure 2.

The first iteration of the noisy student training steps yielded a 6% improvement for the validation loss over Ghana. The second iteration yielded a 1.5% improvement, after which iterations were stopped.

The multisensor deep learning model is based on a modified Inception-ResNet approach (Szegedy et al., 2017) with multiple repeated inception blocks and a total of 13.4 million parameters. Mean squared error loss (MSE) was chosen as it optimizes toward the mean, which is beneficial when many tiles contain few to no structures (Hyndman & Athanasopoulos, 2018). While Fibæk, Keßler, et al. (2021) suggest using stepwise learning rate decay—modifying the implementation to iteratively increase the batch size while keeping the learning rate constant consistently gave better results, which corresponds to the findings of Smith et al., 2018. The area model was modified by clipping the output to ensure values always fall between 0 and 100 and adding batch normalization

TABLE 2 Sentinel products used. Data collected from Sentinel 1A, 1B, 2A, and 2B

Source	Data product	Band/pol	Pixel size
Sentinel 1	GRD	VV	10 m
"	"	VH	"
Sentinel 2	2A	Band 2	"
"	"	Band 3	"
"	"	Band 4	"
"	"	Band 8	"
"	"	Band 5	20 m
"	"	Band 6	"
"	"	Band 7	"
"	"	Band 11	"
"	"	Band 12	"

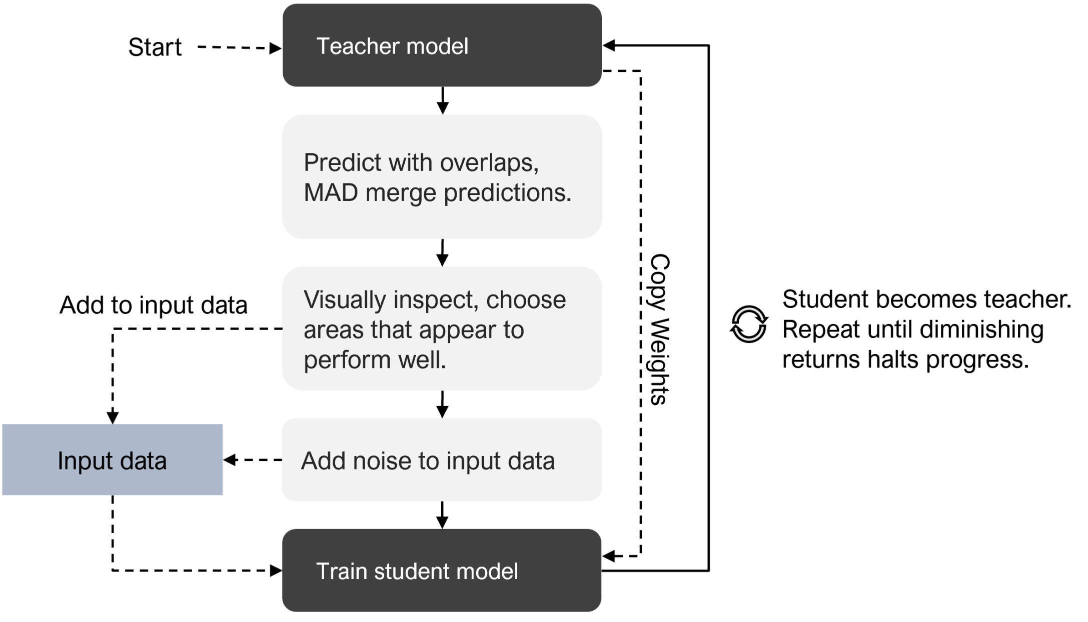

FIGURE 2 Semi-supervised noisy student training steps.

and 2D spatial dropout before each concatenate block. The final model design is available in the data availability statement.

Eleven predictions from distinct overlaps were made for each pixel. These predictions were merged using "Median Absolute Deviation (MAD) Interval Merging" with the naïve implementation being:

1. Select the value with the most adjacent values within one median absolute deviation (MAD) as the Center
2. Take all the values within one MAD of the center, and create weights based on each value's own number of adjacent values within one MAD
3. Take the weighted median of all the values in step 2

Changing the algorithm for merging predictions from average to the median, to mad interval merging leads to reduced noise in the final classification at the cost of increased post-processing time. An implementation of mad interval merging is available in the Buteo toolbox, and an example comparison is shown in Figure 3.

4.1 | Structure type

Four classes were used to segment the structure types. 1. "Residential", 2. "Residential (self-organized)," 3. "Nonresidential," and 4. "No structure." The self-organized class was included to test the separability of the classes and the usefulness of the model in slum mapping. After training the model to extract structural area, the model was frozen except for the last inception block and the tail layers. The unfrozen layers were then retrained, and the activation function changed to SoftMax. The model design is shown in Figure 4.

The ground truth data for structural classes was based on Owusu et al. (2021) accessed through slummap.net, which served as the basis for collecting the ground truth classes for Ghana. The data were digitized using the newest available imagery from Google Earth and Microsoft Bing. Structures that were highly likely to be nonresidential, such as gas stations and marketplaces were classed as such – but mixed-use was not classified in Ghana.

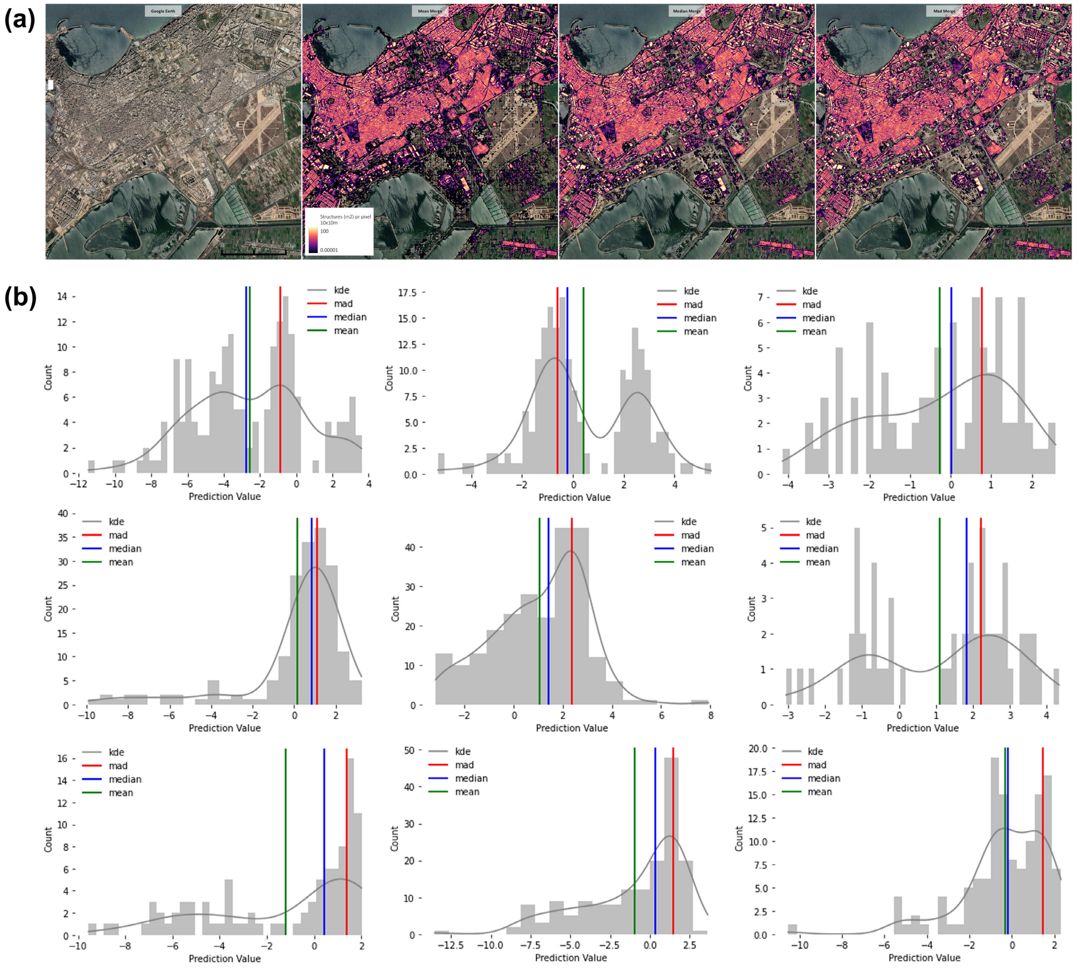

FIGURE 3 Comparison of different methods of merging predictions. mad refers to the MAD interval merging method. Top: Area predictions for eastern Alexandria, Egypt. Bottom: Comparison of prediction merging for sample distributions of three skewed normal distributions of random size. Mad interval merge produces less noise than mean and median merging. The effect is more apparent at the full resolution of the figure.

There is significant room for improvement in ground truth collection for the classes. With the full release of the 2021 Ghana Census, it might be possible to improve the dataset considerably. While some structure type information was available in Egypt from CAPMAS (CAPMAS, 2017), no structural type classification was done there—the population predictions for Egypt are based on an unweighted approach. The approach presented for Ghana is likely applicable to Egypt as well.

4.2 | Population

Tatem (2017) and Tiecke et al. (2017) have shown that it is possible to generate population estimates using structural footprints, census data, and UN population forecasts. The two methods used very high-resolution satellite imagery to extract the building footprints, whereas this study uses Sentinel data and structural area and type instead of structure count. As people move throughout the day population is not static. When people leave

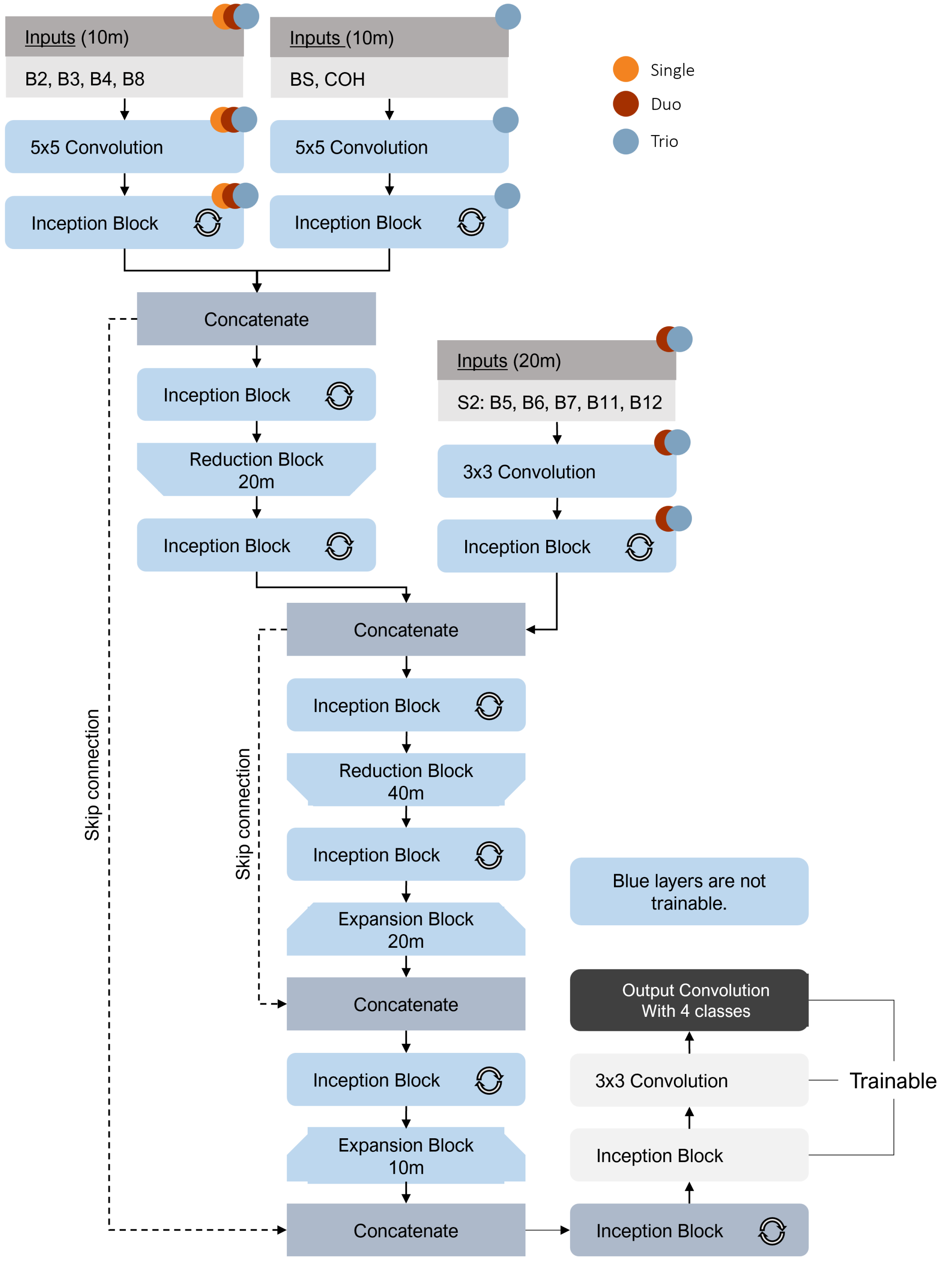

FIGURE 4 Design of the transfer learning structural classification model.

for work during the daytime, the suburbs will be less densely populated than in the mornings and the evenings (Foley, 1954). This study generates daytime and nighttime population estimates by combining structural area estimates with structure classifications. Census data for population counts are often only available at the regional

TABLE 3 Initial weights used for daytime/nighttime estimates

Type	Residential	Residential self-organized	Nonresidential
Unweighted	1.0	1.0	1.0
Nighttime	1.25	1.35	0.4
Daytime	1.05	1.15	0.8

level, and as censuses are usually only conducted every 10 years, it is necessary to use population projections to estimate the population. The population estimates were calculated as illustrated in Equations (1)–(4) from the census, structure type, and area predictions.

$$c_{s_{ij}} = \left(c_{1_{ij}} + c_{2_{ij}} + c_{3_{ij}}\right) + \varepsilon_{fp} \tag{1}$$

$$pc_{x_{ij}} = \frac{c_{x_{ij}}}{c_s} * r_{ij} * w_x \tag{2}$$

$$r_{ij} = \frac{a_{ij}}{p_{ij}} \tag{2.1}$$

$$pop_{tu} = \sum pc_1 + \sum pc_2 + \sum pc_3 \tag{3}$$

$$pop_t = \frac{tp}{pop_{tu}} * pop_{tu} \tag{4}$$

where $c_{1_{ij}}$ is the probability of pixel *i,j* belonging to class 1; $c_{x_{ij}}$ is the probability of pixel *i,j* belonging to class *x*; $c_{s_{ij}}$ is the sum of the class probabilities not "no structure"; ε_{fp} is the floating-point epsilon; p_{ij} is the structure in m^2 pr. person; a_{ij} is the structure coverage predicted for the pixel in m^2; $r_{ij}r$ is the ratio between area in m^2 and population per m^2 in the region; $pc_{x\,ij}$ is the unscaled population of the class; w_x is the weight given to class *x*; pop_{tu} is the unscaled weighted population of the raster; and pop_t is the scaled weighted population of the raster.

The square meter of structure per person was calculated using the 2021 census for Ghana and the 2017 census for Egypt (CAPMAS, 2017; Statistical Service Ghana, 2021), and the weights were set as shown in Table 3.

The weights were chosen to serve as a conservative baseline for further investigation. Future studies should investigate the relationship between different urban structures and temporal populations. The borders between the estimates of structural area pr. person for each smallest available census zone were feathered using the average value of a 2.5 km circular buffer to create a smoother transition and then scaled back to the original sum.

4.3 | Model training and requirements

The area extraction model was trained on a Dell XPS 157590 computer with the following specifications:

CPU	Intel i9 9980HK 8/16 cores
RAM	64GB
OS	Windows 11
GPU	Nvidia GeForce RTX 3090 - 24GB Ram (External)

The model uses 16-bit floating-point operations, and at least 24 GB GPU Ram is required to train the model. The training time was approximately 50 hours for the initial area extraction model. It is possible to make model predictions without a GPU, but at least 8 GB of GPU RAM is recommended to make predictions for areas larger than a city block.

5 | RESULTS

The following section shows the results from the three predictions of area, type of structure and the derived population estimates named SenPop. All predictions were made using 10 additional offset overlaps and merged using MAD interval merging.

5.1 | Structural area

Comparisons to Google Open Buildings (Sirko et al., 2021) were made for the predicted area per pixel to assess the relative accuracy of the structural area predictions. Binary comparisons were made to the Global Human Settlement Layer S2-Net (Corbane et al., 2020), OpenStreetMap (Keßler, 2015), and ESA World Cover (ESA, 2020). A comparison of area per pixel accuracy for Uganda and Tanzania could likewise be made to the Microsoft Open Buildings dataset (Microsoft, 2021); however, only the target study areas were evaluated.

Sixteen test sites were distributed across Ghana, covering 94.5 $km^{2,}$ not overlapping any training, test or student sites used in training the models and in Egypt, four test sites were used. The TPE accuracy metric is defined as the ratio between the predictions' sum and the labels' sum expressed in per cent—with 100% being the ideal value. A default threshold of 0.5 m^2 of structures was used for Sentinel and Google Predictions for the binary comparisons.

Tables 4–7 show accuracy comparisons for the structural area predictions over Ghana and Mediterranean Egypt at 10 × 10 m and 100 × 100 m resolution. Figure 5 shows an example prediction for Accra—Ghana, and compares them with the ground truth data and the Google Open Buildings at different precision intervals. Figures 6 and 7 show the area predictions over Mediterranean Egypt and compare them with Google Earth Imagery. Figure 8 compares area predictions when resampled to 100 × 100 m resolution using the sum method. rTPE and bTPE refer to the Total Percentage Error for the regression and binary data. The total percentage error is the sum of the predicted values divided by the sum of the label values expressed as a percentage.

There is a clear tendency to create smooth predictions using the Sentinel-based area predictions. The Google Open Buildings 50% layer performs better or on par with the Sentinel predictions for most metrics. The Google Open Buildings is based on much higher resolution input data, which explains its tendency to produce sharper predictions. However, the Sentinel predictions model outperforms the other models in estimating the total area of structures present. The over smoothing of the Sentinel predictions results can be alleviated by resampling to a lower resolution, as shown in Table 5. The binary classifications were resampled by the mode, whereas the other products were resampled by the average.

When investigating the resampled data, the Sentinel-based predictions outperform the Google estimates at all precision intervals for area predictions. The binary predictions are close—with the Sentinel predictions performing better when comparing the total predicted sums. Having accurate sums enables reliable predictions of population using the structural area predictions.

The comparison between the Sentinel predictions and the Google products is only provided to highlight relative accuracy. Parts of the Google Open Buildings Dataset, along with Open Street Map, were incorporated into the creation process of the ground truth datasets, which means that the accuracy metrics for OpenStreetMap and Google Open Buildings might be overestimated.

TABLE 4 Comparison of area predictions in Ghana—10 m pixel size

Source	Pixel size (m)	MAE	RMSE	rTPE	Bal. ACC	ACC	F1	Precision	Recall	bTPE (%)
Sentinel predictions (SP)	**10**	**3.664**	**10.596**	**100.5%**	**0.931**	**0.903**	**0.745**	**0.604**	**0.971**	**154.7**
SP 1 m^2 binary threshold	10	–	–	–	0.934	0.915	0.775	0.649	0.961	148.2
SP 5 m^2 binary threshold	10	–	–	–	0.918	0.941	0.82	0.766	0.884	115.4
Google OB 50%	10	1.495	8.014	76.7%	0.935	0.979	0.923	0.977	0.874	86.2
Google OB 80%	10	1.861	9.212	68.7%	0.894	0.967	0.876	0.982	0.791	77.6
Google OB 85%	10	2.261	10.478	60.2%	0.851	0.955	0.822	0.986	0.704	68.8
Google OB 90%	10	2.779	11.945	49.5%	0.796	0.940	0.741	0.988	0.593	57.8
OpenStreetMap	10	4.042	14.696	26.1%	0.622	0.889	0.392	0.966	0.246	24.5
ESA WorldCover	10	–	–	–	0.895	0.914	0.747	0.656	0.868	127.6
GHS S2-NET $p > 0.0$	10	–	–	–	0.881	0.810	0.601	0.433	0.982	218.5
GHS S2-NET $p > 0.5$	10	–	–	–	0.695	0.888	0.524	0.694	0.421	58.4

Note: Binary comparisons in light gray. Google OB refer to Google open buildings. The bolded row is the predictions produced by the methodology described in this article.

TABLE 5 Comparison of area predictions in Ghana—100 m pixel size

Source	Pixel size (m)	MAE	RMSE	rTPE (%)	Bal. ACC	ACC	F1	Precision	Recall	bTPE
Sentinel predictions	**100**	**0.965**	**2.344**	**99.97**	**0.965**	**0.967**	**0.950**	**0.942**	**0.959**	**93.7**
Google OB 50%	100	1.214	3.474	76.8	0.978	0.985	0.977	0.995	0.959	88.7
Google OB 80%	100	1.618	4.519	68.7	0.975	0.983	0.974	0.998	0.951	87.8
Google OB 85%	100	2.051	5.624	60.2	0.970	0.979	0.968	0.998	0.940	86.8
Google OB 90%	100	2.605	6.820	49.4	0.959	0.973	0.957	0.998	0.920	84.8
OpenStreetMap	100	3.799	9.191	26.2	0.674	0.783	0.516	0.995	0.349	32.3
ESA WorldCover	100	–	–		0.893	0.903	0.856	0.847	0.864	84.0
GHS S2-NET p > 0.0	100	–	–		0.791	0.730	0.705	0.553	0.973	152.1
GHS S2-NET p > 0.5	100	–	–		0.558	0.689	0.266	0.612	0.170	15.7

Note: Binary comparisons in light gray. The bolded row is the predictions produced by the methodology described in this article.

TABLE 6 Comparison of area predictions in Egypt – 10 m pixel size

Source	Pixel size (m)	MAE	RMSE	rTPE	Bal. ACC	ACC	F1	Precision	Recall	bTPE (%)
Sentinel predictions	**10**	**3.533**	**11.047**	**93.7%**	**0.948**	**0.929**	**0.777**	**0.647**	**0.973**	**150.3**
Google OB 50%	10	4.070	14.795	59.3%	0.881	0.955	0.814	0.848	0.783	92.3
Google OB 80%	10	5.640	20.005	15.1%	0.620	0.901	0.384	0.895	0.245	27.4
Google OB 85%	10	5.812	20.454	10.7%	0.585	0.893	0.289	0.915	0.172	18.8
Google OB 90%	10	5.928	20.741	7.8%	0.561	0.888	0.217	0.943	0.123	13.0
OpenStreetMap	10	6.366	21.404	0.01%	0.500	0.873	0.000	1.000	0.000	0.01
ESA WorldCover	10	–	–	–	0.901	0.891	0.681	0.542	0.914	168.7
GHS S2-NET p > 0.0	10	–	–	–	0.721	0.514	0.343	0.207	0.999	483.0
GHS S2-NET p > 0.5	10	–	–	–	0.86	0.904	0.679	0.588	0.802	136.5

Note: Binary comparisons in light gray. The bolded row is the predictions produced by the methodology described in this article.

TABLE 7 Comparison of area predictions in Egypt—100 m pixel size

Source	Pixel size (m)	MAE	RMSE	rTPE	Bal. ACC	ACC	F1	Precision	Recall	bTPE (%)
Sentinel predictions	**100**	**1.341**	**3.366**	**94.0%**	**0.959**	**0.960**	**0.933**	**0.912**	**0.956**	**104.9**
Google OB 50%	100	2.844	7.975	57.9%	0.959	0.967	0.943	0.946	0.941	99.5
Google OB 80%	100	5.301	14.114	14.6%	0.841	0.903	0.807	0.971	0.691	71.1
Google OB 85%	100	5.542	14.557	10.3%	0.765	0.859	0.691	0.981	0.534	54.4
Google OB 90%	100	5.710	14.845	7.4%	0.657	0.798	0.478	1.000	0.314	31.4
OpenStreetMap	100	6.156	15.187	0.01%	0.500	0.705	0.000	1.000	0.000	0.01
ESA WorldCover	100	–	–	–	0.917	0.904	0.854	0.776	0.948	122.2
GHS S2-NET p > 0.0	100	–	–	–	0.575	0.400	0.496	0.330	1.000	303.2
GHS S2-NET p > 0.5	100	–	–	–	0.750	0.852	0.667	1.000	0.500	50.0

Note: Binary comparisons in light gray. The bolded row is the predictions produced by the methodology described in this article.

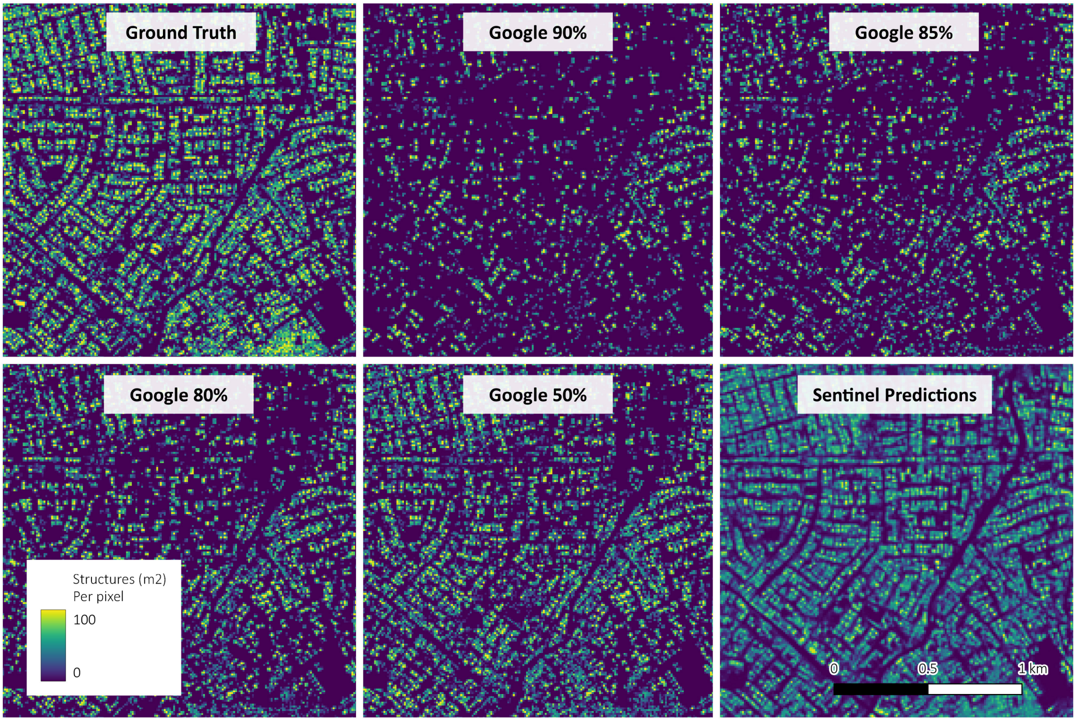

FIGURE 5 Comparison between Google open buildings and the sentinel extracted buildings. Accra—West Legon.

The Sentinel predictions for Egypt's Mediterranean Coast perform similarly to Ghana but perform better relative to the comparison datasets. Egypt has many structures in very close proximity, which are difficult to separate for the model. Furthermore, roads and the sides of canals are sometimes misclassified as structures—which happens less in Ghana. All the models significantly underestimate the total area of structures in Egypt. The reason seems to be underestimation in dense urban cores.

At the 100 m scale, the relative accuracy of the Sentinel Predictions increases, as seen in the Ghana case. While the model underestimates the total area per tile, it does better than other state-of-the-art methods.

Visually inspecting the results shows the tendency to create too "smooth" results using the Sentinel-based predictions at 10 × 10 m resolution. While the patterns and buildings are visible—the individual structures are awarded too low an area value. It does well at mapping self-organized areas where the Google Open Buildings layers have trouble, as seen in the bottom right corner of Figure 11. In rural areas, the same pattern is visible, in that a similar count of structures is predicted, but the predictions are less sharp.

In Mediterranean Egypt, there is also a tendency to over smooth the results at the native 10 × 10 m resolution of the predictions. Some roads are misclassified as structures, especially if the road has sharp embankments. The model does well at expressing the location and general density of structures, as shown in the flat coastal areas along the Nile Delta Estuaries in Egypt. Inland, in the rocky, hilly areas, the model performs worse, sometimes confusing rising rock faces with structures. This is likely due to the double bounce effect on the Sentinel 1 satellites (Koppel et al., 2017).

Comparing the results at the resampled 100 × 100 m resolution show that the performance of the Sentinel predictions increases when the tendency to over smooth is alleviated by resampling. The patterns and total sum of area are generally very close to the patterns expressed by the ground truth data.

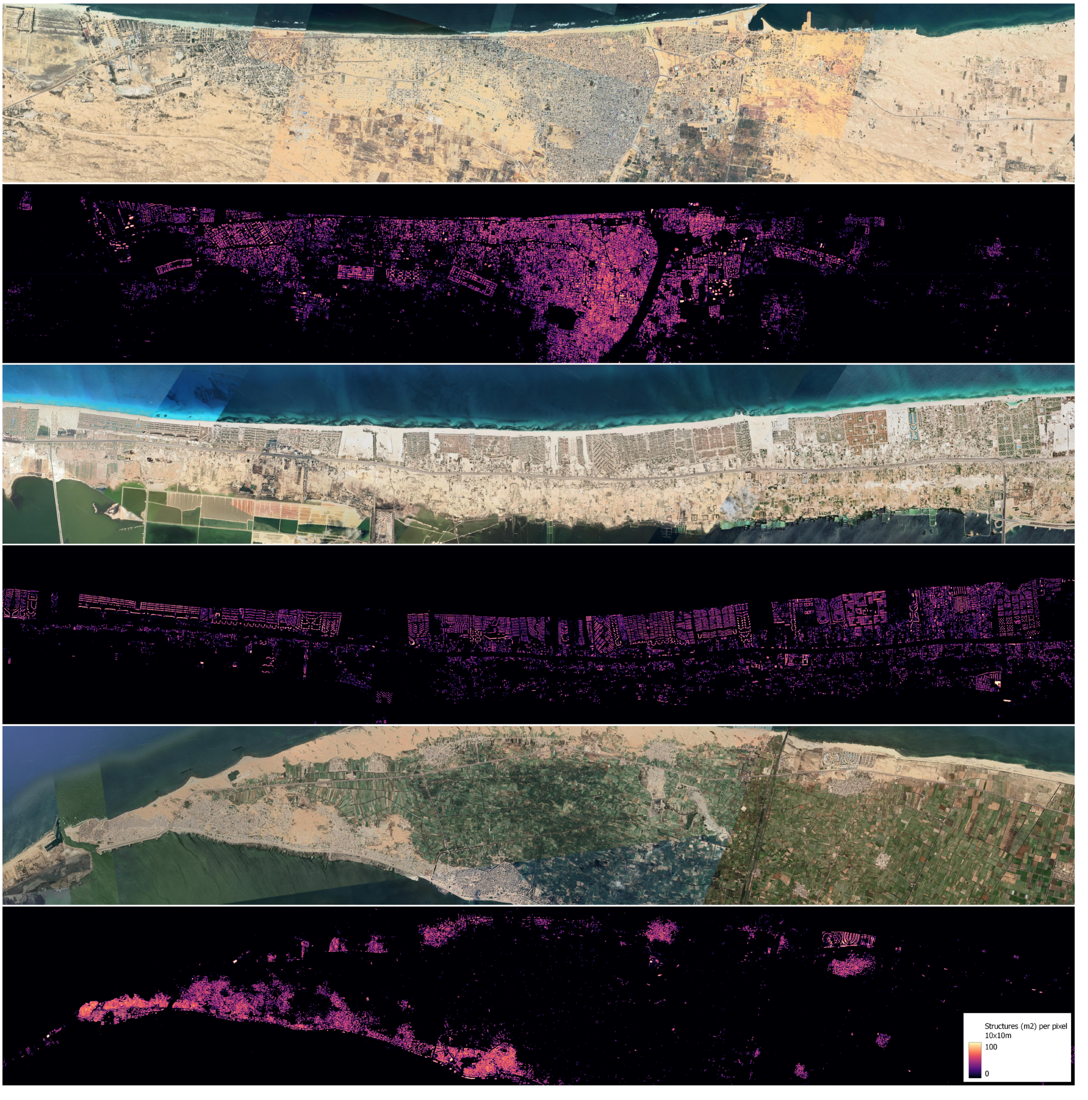

FIGURE 6 Comparison of sentinel area predictions and Google earth along the Mediterranean coast of Egypt.

An important distinction between the methodologies is that the Minimum mapping unit of the Sentinel prediction model is 10 by 10 m. In contrast, Google Open Buildings provides vectors derived from very high-resolution imagery.

Mapping the area of structures instead of binary classification of built-up areas allows better comparison of different types of urbanization over time. It enables the estimations of locations where cities are getting denser and where urban sprawl is increasing—a distinction key to many studies such as Moran et al. (2018).

5.2 | Structure classification

A classification of structures was made to create a translation layer between daytime and nighttime population layers based on the change in population patterns throughout the day and year (Foley, 1954). While the area, or ideally

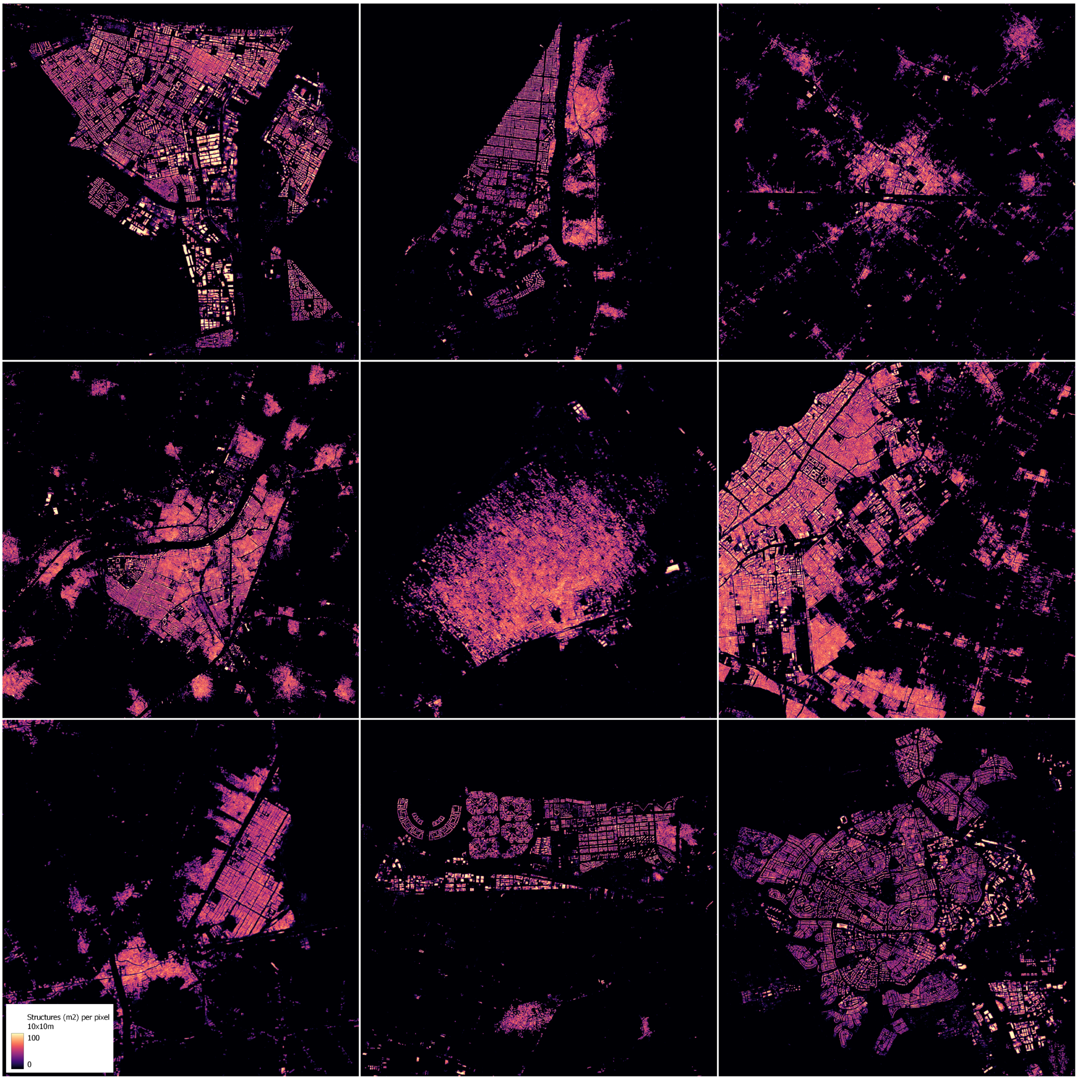

FIGURE 7 Example sentinel area prediction for cities and villages in the Nile Delta and along the Mediterranean coast, Egypt.

the volume of structures, and census data or forecasts can be used to estimate population, they do not consider the temporal aspects of population estimation (Tatem, 2017). Using the model trained to extract structural area and deep transfer learning to classify structures can enhance the population estimates. The classes were balanced by taking the majority class of each 32×32 tile in the training sites and oversampling underrepresented classes to reach 10,000 tiles for each class. In Ghana, Google Earth and Microsoft Bing imagery was used to create the training data. Figure 9 shows the confusion matrix for the classification of 6500 separate test tiles in Ghana. While

The labels consisted of the overlap of structures belonging to each class within the pixel. As such, each label was a $32 \times 32 \times 4$ matrix, with the last four values corresponding to the ratio of the pixel covered by the given class from zero to one, with the values totalling one. Categorical cross-entropy was used as the loss function. The model generally predicted the majority class well, with "uninhabited/no structures" often being mistaken. Equation (1) redistributes the probability of class 1 to the other three classes, reducing the impact of this misclassification. Table 8 shows an example pixel, where the structural area prediction model predicted $24\,m^2$ of structures and the

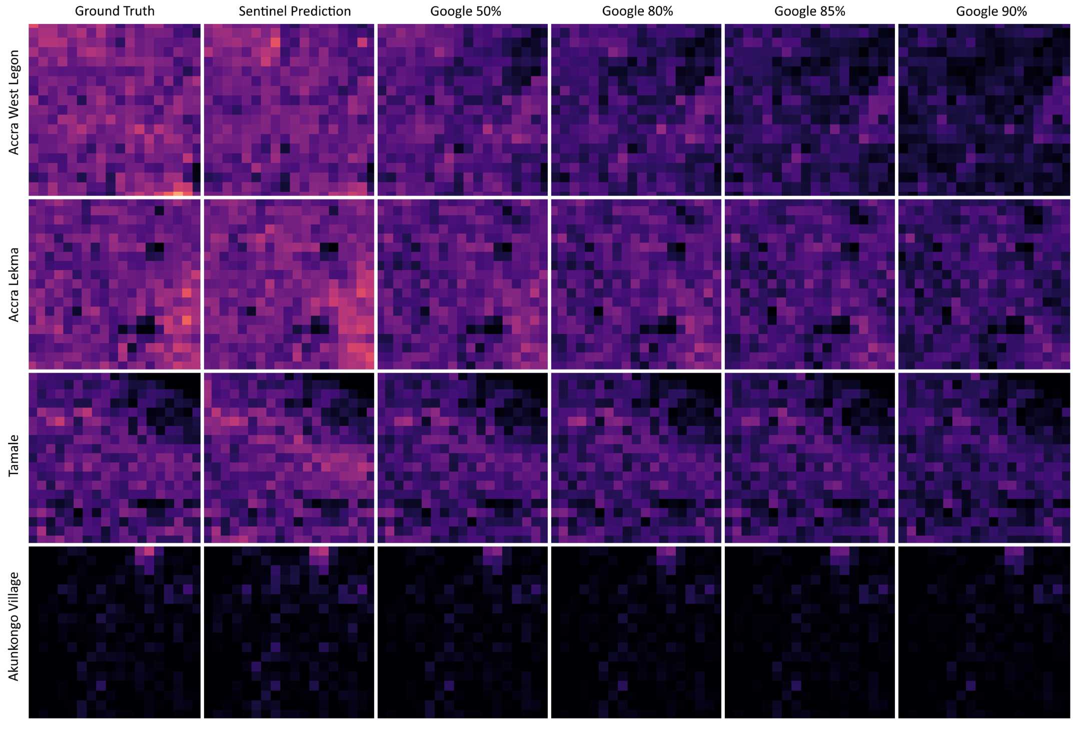

FIGURE 8 Comparing the results at the 100 by 100 m per pixel resolution Accra, Ghana.

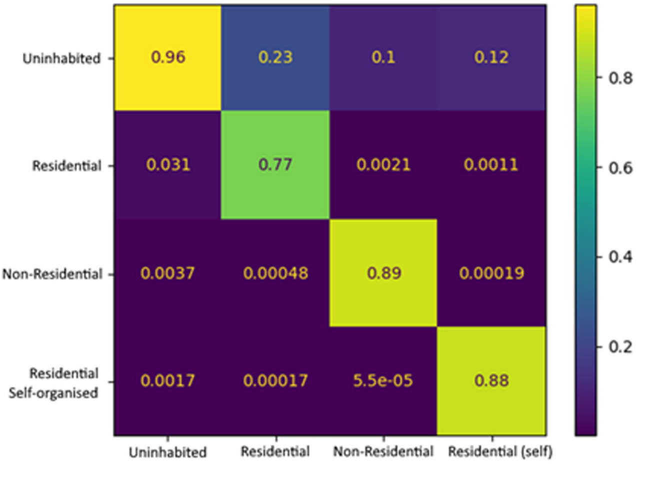

FIGURE 9 Confusion matrix of the structural classifications in Ghana.

TABLE 8 Example processing of pixel from structural classification to area type per pixel

Type	Uninhabited	Residential	Nonresidential	Residential (self)
Probability	50%	35%	5%	10%
Weight	0.0%	100%	100%	100%
Probability (adj)	0%	70%	10%	20%
Area of type	0.0 * 24 = 0 m^2	0.7 * 24 = 16.8 m^2	0.1 * 24 = 2.4 m^2	0.2 * 24 = 4.8 m^2

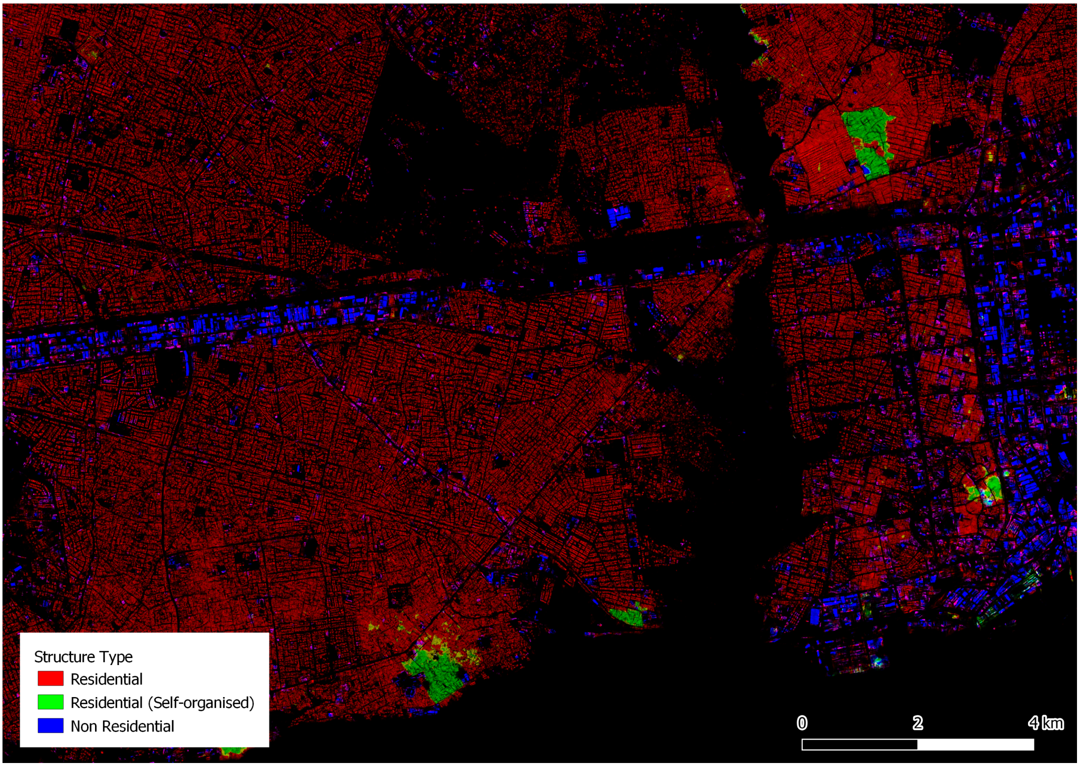

FIGURE 10 Accra East & Tema, Ghana. Structure type. RGB render. Black is uninhabited.

probabilities were distributed as shown in Table 8. Note that while the highest probability is uninhabited, some residential area is still added to the pixel.

Figure 10 show an RGB render of the predicted classes containing structures.

5.3 | Population (SenPop)

Population estimates were made by combining the UN forecasts with the census data from 2021 for Ghana and 2017 for Egypt. The smallest population zones available for the authors was the regional level.

Figure 11 shows the population predictions in Accra, Ghana and the difference between estimated daytime and nighttime populations as described in Section 4.2. Tables 9 and 10 and Figures 12 and 13 show the relative population predictions intercompared with WorldPop and Meta estimates. Figures 14 and 15 visualize the differences between the different population predictions.

6 | DISCUSSION

Globally, there is an increasing amount of training data available for structures and efforts to make large deep learning models accessible through cloud computing, knowledge distillation, and public training datasets (Hamed Alemohammad, 2019). This means that methods such as the ones described in this study are becoming increasingly accessible. The population predictions presented in this study show that a Sentinel-based approach to population estimation is feasible for Africa, Western Asia, and Europe and provides a baseline for future studies.

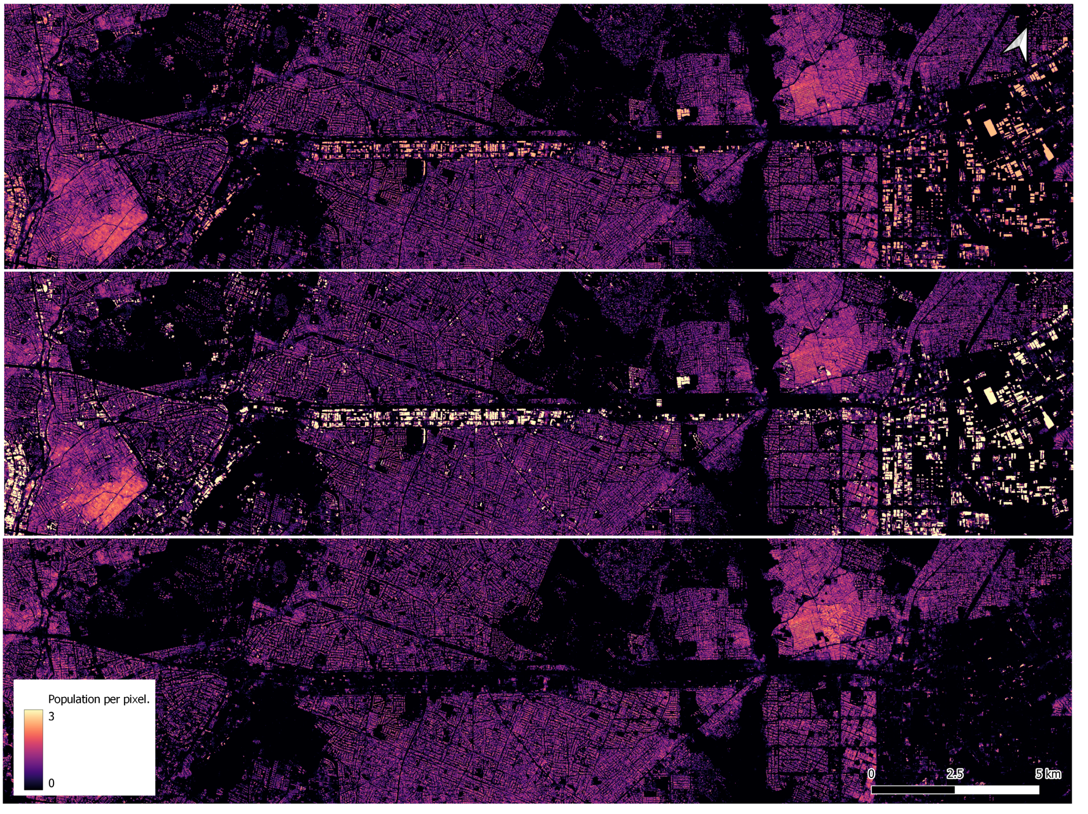

FIGURE 11 Accra—Tema industrial belt. Comparison of population scaled by daytime. (Top) Unweighted, (middle) daytime, and (bottom) nighttime.

The population predictions have a higher spatial resolution than currently offered globally, and the reliance on Sentinel data means that the potential temporal resolution is high. As it is only possible to do an intercomparison of population estimates, assessing the absolute accuracy of the estimates is not possible without collecting additional ground truth data. The produced population estimates differ from the current state-of-the-art, both by relying on significantly lower resolution data and building on the area of structures instead of the count (Tatem, 2017). The addition of structural classifications allows the estimation of daytime and nighttime population, which furthers the granularity offered by the models. An interesting topic for further research is temporary populations, such as traffic hotspots and commuting routes. Such estimates could be made using extracted road networks from Sentinel imagery and points of interest data from OpenStreetMap.

Fibæk, Keßler, et al. (2021) and Frantz et al. (2021) show that it is possible to estimate the building height and volumes of structures in Northern Europe using Sentinel 1 and 2 imagery. Using the volume instead of the area for the population estimates would increase the robustness of the predictions, especially in dense urban areas and clusters with industrial or high-rise buildings. While training data on the structural volume is available for Denmark, no such data was available for the case study areas. A far-field application of the Danish model to Ghana was attempted but did not yield promising results.

Relying on Sentinel data means the proposed methodology has a high temporal resolution. For many areas in Africa, Sentinel data are only available from one orbital direction. Fibæk, Keßler, and Arsanjani (2021) investigated the effects of this on the accuracy of structural predictions and found a significant decrease in accuracy. If both orbital directions become available, it will likely be possible to reduce the number of errors caused by the alignment of structures and natural features (Koppel et al., 2017). Introducing data from a terrain model, such as the

TABLE 9 Comparison of population predictions Ghana—People per location normalized

Location	UW	Night (%)	Day (%)	WP (%)	WP-C (%)	WP UN (%)	WP UN-C (%)	Meta (%)
1. Aburaso	17,664	103.5	96.3	30.7	64.3	28.9	60.5	74.5
2. Agogo	44,887	101.0	98.9	18.0	50.3	16.9	47.3	55.8
3. Agona Swedru	125,585	103.0	96.9	57.4	88.1	54.0	82.9	120.4
4. Ashaiman	63,163	104.4	100.5	69.9	92.9	65.7	87.4	33.5
5. Bolgatanga Center	11,110	90.7	109.5	47.0	51.6	44.2	48.6	76.2
6. Buya	4553	103.1	96.7	15.9	75.0	15.0	70.6	142.2
7. Ejura	40,330	101.2	98.7	18.4	47.3	17.3	44.5	89.8
8. Fort Orange	41,105	95.6	105.0	216.2	202.7	203.3	190.6	156.9
9. Gbadzeme	1212	102.4	97.4	45.6	111.7	42.9	105.0	124.6
10. Kotei	76,554	102.4	97.4	188.2	239.4	177.0	225.2	551.4
11. Mamekrobo	11,190	101.7	98.3	24.5	93.2	23.0	87.7	129.3
12. Olemuni	1227	104.0	95.8	19.1	95.9	17.9	90.2	252.4
13. Sakasaka	27,556	100.6	99.2	40.9	36.4	38.5	34.2	26.6
14. Tema & Industrial	48,234	62.9	137.4	80.8	78.4	76.0	73.7	29.6
15. Tinga	12,315	103.7	96.1	7.2	34.1	6.8	32.1	45.5
16. Wa	170,367	101.5	98.3	26.9	37.6	25.3	35.4	53.6
17. Wulugu	8831	102.8	97.1	21.8	71.1	20.5	66.8	128.6
18. Yeji	67,536	102.5	97.4	17.5	57.3	16.5	53.9	60.2
	773,421	99.3	101.0	63.4	86.9	59.6	81.7	121.3

Note: Unweighted SenPop predictions were used as the baseline.
Abbreviations: C, constrained; UN, UN-adjusted; UW, unweighted; WP, WorldPop.

TABLE 10 Comparison of population predictions Egypt—People per location normalized

Location	UW	WP (%)	WP-C (%)	WP UN (%)	WP UN-C (%)	Meta (%)
1. Dugmayrah	19,359	61.8	68.9	65.7	73.3	78.7
2. Tanta Metro	677,648	99.9	107.7	106.4	114.6	91.1
3. El-Agamy	509,736	102.1	102.8	108.7	109.4	88.9
4. El Dabaa	72,759	24.5	27.6	26.1	29.3	15.0
5. El Manzala & Suburbs	242,245	88.6	94.2	94.3	100.3	78.2
6. Bir El-Abd	87,513	25.9	22.6	27.6	24.1	35.6
7. Baltim Summar	36,663	94.6	94.2	100.7	100.3	132.2
8. Izbat Villages	47,412	106.5	116.4	113.4	123.9	110.9
9. Ad Dilinjat Agriculture	26,008	65.8	88.5	70.1	94.3	80.3
10. Desouk	228,030	101.2	111.6	107.8	118.8	86.9
	1,947,374	92.3	97.7	98.3	104.0	84.1

Note: Unweighted SenPop predictions were used as the baseline.
Abbreviations: C, constrained; UN, UN-adjusted; UW, unweighted; WP, WorldPop.

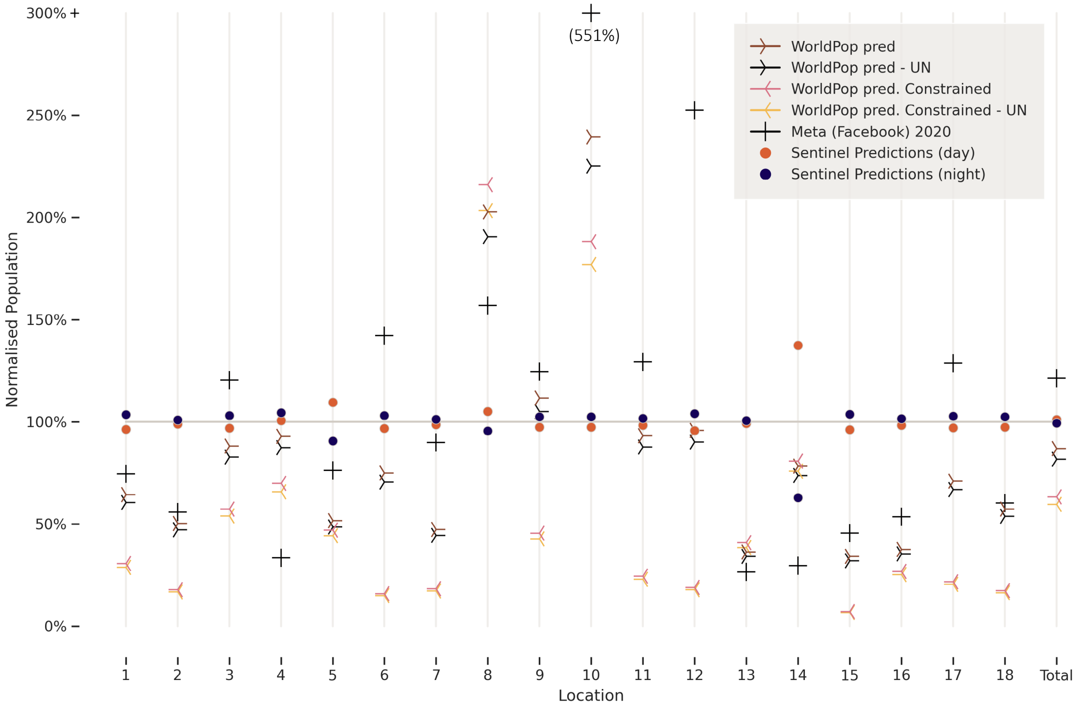

FIGURE 12 Intercomparison of population estimates in Ghana. Normalized to the unweighted sentinel predictions. Locations are defined in Table 9.

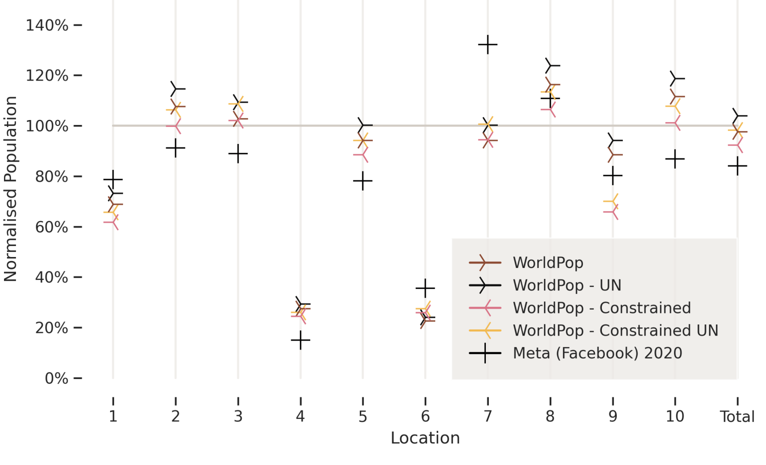

FIGURE 13 Intercomparison of population estimates in Egypt. Normalized to the unweighted sentinel predictions. Locations are defined in Table 10.

Copernicus Global DEM, into the deep learning model itself might further reduce the errors caused by the single orbital direction. The Sentinel 1 data used in the study was processed with the SRTM DEM, which should also be updated to use the Copernicus Global DEM.

The described methodology is trained and produces predictions based on single timestamps from merged mosaics: Changing the method to work on time series might further increase the accuracy of the estimates,

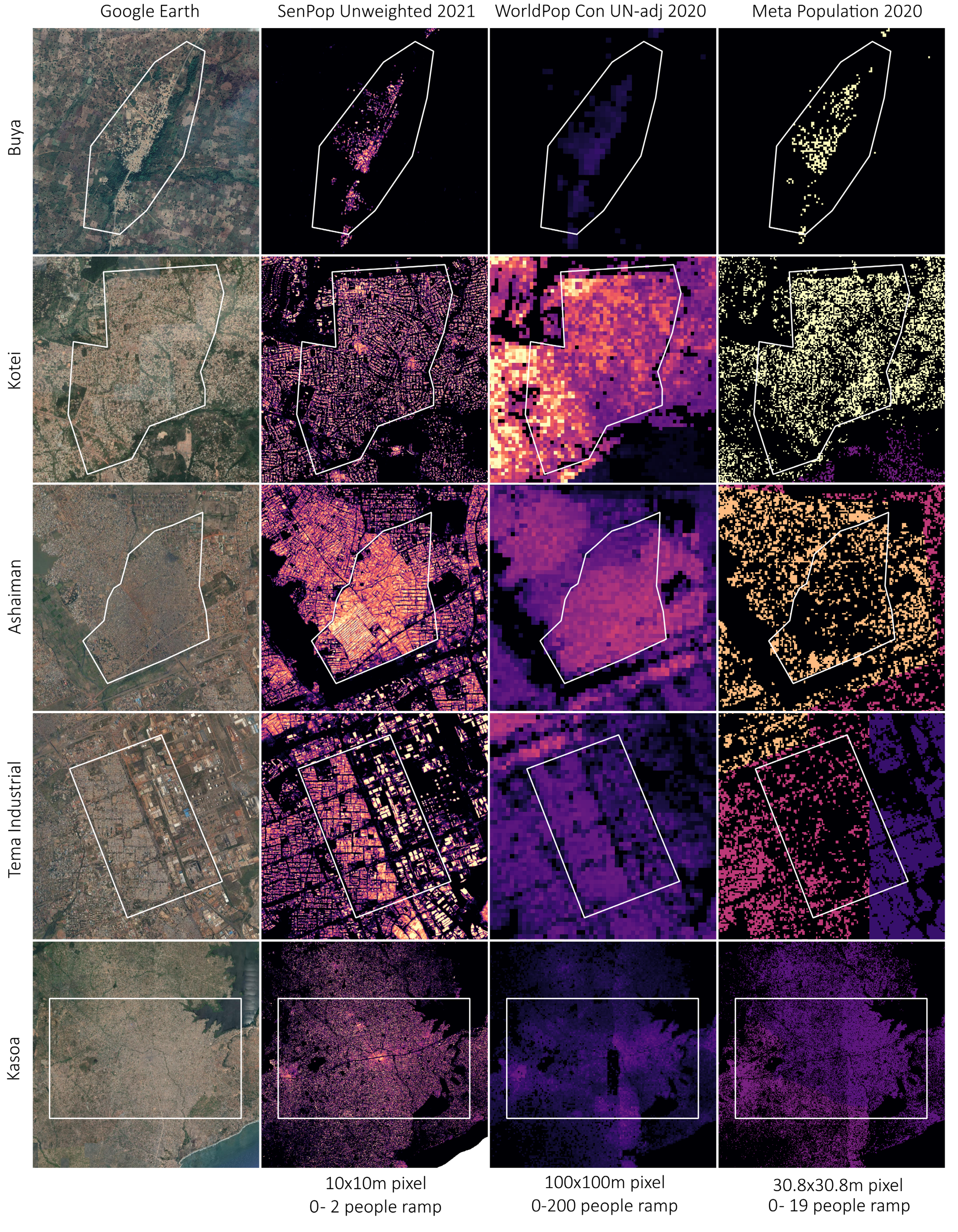

FIGURE 14 Visual intercomparison of population estimates at their native resolutions. Cities, suburbs, and villages in Ghana.

as might including metadata from the image capture into the model. Updating the model design to incorporate Partial Convolution-Based Padding would likely reduce the post-processing time required by the overlaps and mad interval merging approach by reducing or removing the border noise produced by the zero-padding

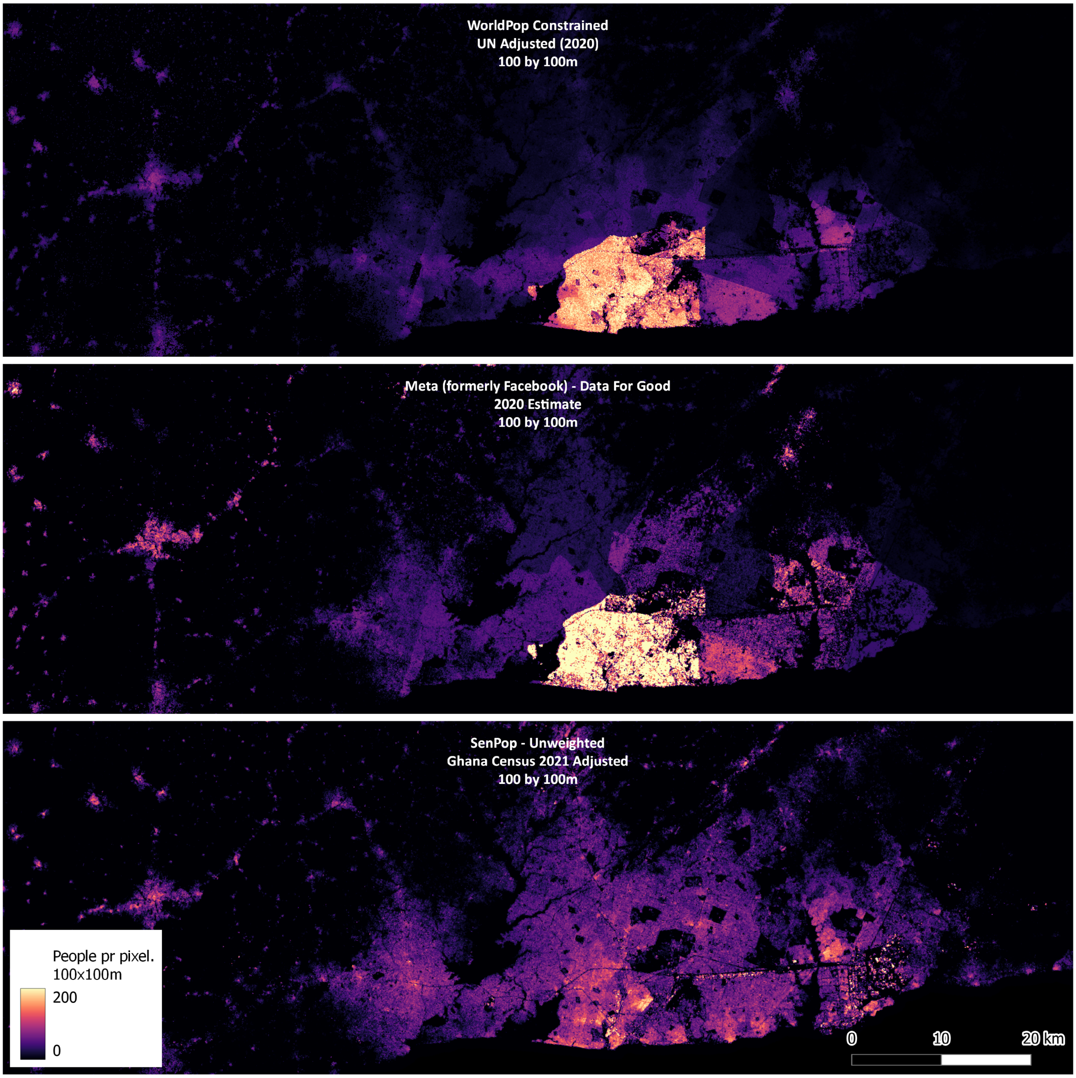

FIGURE 15 Intercomparison between the WorldPop constrained UN-adjusted layer, meta population predictions, and SenPop. SenPop and meta predictions were resampled to 100 × 100 m for comparison. Accra, Ghana.

convolution boundaries (Liu et al., 2018). Creating additional synthetic Sentinel imagery for training using Generative Adversarial Networks could ensure the robustness of the model across multiple seasons without the need for collecting a large amount of input data (Mohandoss et al., 2020).

Significantly less labeled data were created for Ghana (47 km^2) and Egypt (9 km^2) than what was used to train the initial Danish models (42,933 km^2 over two seasons; see Table 1). The amount of training data required to use the proposed methodology in new areas would depend on the geographical likeness to the areas included within the study. No additional training data may be necessary for highly similar regions, like northern Germany to Denmark or Togo to Ghana. Besides the singular peak of Mount Kilimanjaro, none of the study areas have covered mountainous terrains. Mountainous terrain would significantly impact shadows, and radar backscatter—including a modern terrain model into the model design could improve its ability to generalize to these areas. Without including DEM data, significant additional training data are likely needed to apply the methodology to such geographies. Only 9 km^2 of labels were gathered to create the predictions for Egypt's Mediterranean Coast, which can serve as a tentative lower bound of the required local data necessary for successful predictions.

7 | CONCLUSION

Our findings show that it is possible to create population predictions from Sentinel data. The models, trained predominantly on Danish data, work well once fitted with minor amounts of local data. The area predictions compare well with state-of-the-art methods that use very high-resolution satellite imagery, although with a minimum mapping unit of 10 m. The area prediction model serves as the foundation of the methodology enabling structural type classifications through transfer learning, which in turn can be converted to population estimates. The intercomparison of population estimates shows that the SenPop prediction exhibits the same spatial tendencies as the global offerings at a higher spatial resolution, good performance in self-organized areas, and state-of-the-art performance at the 100 × 100 m resolution. Sentinel data as the model input data means that the produced models are scalable and globally applicable, and with the availability of annual global Sentinel 1 and 2 mosaics, it is possible to create global maps of the structural area and population estimates using the deep learning methods presented. The area prediction model can serve as a baseline for further studies investigating energy needs, emissions, disaster management, and service levels, which are central to tracking and guiding the progress toward the UN Sustainable Development Goals.

DATA AVAILABILITY STATEMENT

The full set of produced predictions are available at: Fibaek (2022), "Area and Population Estimates from Sentinel Data—Ghana and Mediterranean Egypt", Mendeley Data, V1, doi: 10.17632/gf8v525tm6.1, https://data.mendeley.com/datasets/gf8v525tm6/1. The core code to recreate the study is available at: https://github.com/casperfibaek/population_estimates. The Buteo Toolbox is available at: https://github.com/casperfibaek/buteo.

ACKNOWLEDGMENT

Casper Fibæk has received funding from the Danish Innovation Fund and the NIRAS/Alectia Foundation. We would like to thank Professor Hany Ayad from the University of Alexandria and Foster Mensah from the University of Ghana for assistance in collecting and verifying ground truth data. Open access funding enabled and organized by ProjektDEAL.

CONFLICT OF INTEREST

On behalf of the authors, I, Casper Fibæk, declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Abou-Mahmoud, M. M. E. (2021). Assessing coastal susceptibility to sea-level rise in Alexandria, Egypt. *The Egyptian Journal of Aquatic Research*, *47*(2), 133–141. https://doi.org/10.1016/J.EJAR.2021.04.002

Agency European Space, E. (2021). *Science toolbox exploitation platform*. SNAP Download. http://step.esa.int/main/

CAPMAS. (2017). *Egypt census results 2017* https://www.capmas.gov.eg/party/party.html

Corbane, C., Syrris, V., Sabo, F., Politis, P., Melchiorri, M., Pesaresi, M., Soille, P., & Kemper, T. (2020). Convolutional neural networks for global human settlements mapping from Sentinel-2 satellite imagery. *Neural Computing and Applications*, *33*(12), 6697–6720. https://doi.org/10.1007/S00521-020-05449-7

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., Blondel, V. D., & Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(45), 15888–15893. https://doi.org/10.1073/pnas.1408439111

Dijkstra, L., Florczyk, A. J., Freire, S., Kemper, T., Melchiorri, M., Pesaresi, M., & Schiavina, M. (2021). Applying the degree of urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics*, *125*, 103312. https://doi.org/10.1016/j.jue.2020.103312

Ecopia. (2021). *Digitize Africa*. https://www.ecopiatech.com/africa

Elnabwy, M. T., Elbeltagi, E., El Banna, M. M., Elshikh, M. M. Y., Motawa, I., & Kaloop, M. R. (2020). An approach based on Landsat images for shoreline monitoring to support integrated coastal management—A case study, Ezbet Elborg, Nile Delta, Egypt. *ISPRS International Journal of Geo-Information*, *9*(4), 199. https://doi.org/10.3390/IJGI9040199

ESA. (2020). *WorldCover 2020*. https://worldcover2020.esa.int/

Esri. (2020). *Land cover*. https://livingatlas.arcgis.com/landcover/

Fabbri, K. P. (1998). A methodology for supporting decision making in integrated coastal zone management. *Ocean & Coastal Management*, *39*(1–2), 51–62. https://doi.org/10.1016/S0964-5691(98)00013-1

Fibaek, C. S. (2022). *Buteo*. Github. https://github.com/casperfibaek/buteo

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*, 102628. https://doi.org/10.1016/j.jag.2021.102628

Fibæk, C. S., Laufer, H., Keßler, C., & Jokar Arsanjani, J. (2021). Geodata-driven approaches to financial inclusion—Addressing the challenge of proximity. *International Journal of Applied Earth Observation and Geoinformation*, *99*, 102325. https://doi.org/10.1016/J.JAG.2021.102325

Foley, D. L. (1954). Urban daytime population: A field for demographic-ecological analysis. *Social Forces*, *32*(4), 323–330. https://doi.org/10.2307/2574113

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 Sensing of Environment*, *252*, 112128. https://doi.org/10.1016/j.rse.2020.112128

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

Ghana Statistic Office. (2021). *Ghana 2021 population and housing census, Volume 1—Preliminary report*. https://census2020.statsghana.gov.gh/gssmain/fileUpload/reportthemelist/PRINT_COPY_VERSION_FOUR22ND_SEPT_AT_8_30AM.pdf

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., & Cresson, R. (2017). Orfeo ToolBox: Open source processing of remote sensing images. *Open Geospatial Data, Software and Standards*, *2*(1), 1–8. https://doi.org/10.1186/s40965-017-0031-6

Hamed Alemohammad. (2019). Radiant ML Hub: A Cloud Based Commons for Geospatial Training Datasets. ESIP Summer Meeting. https://pdfs.semanticscholar.org/daee/2836cf84de198cb5c6529327cb3d3f597f6d.pdf

Haberl, H., Wiedenhofer, D., Schug, F., Frantz, D., Virag, D., Plutzar, C., Gruhler, K., Lederer, J., Schiller, G., Fishman, T., Lanau, M., Gattringer, A., Kemper, T., Liu, G., Tanikawa, H., Van Der Linden, S., & Hostert, P. (2021). High-resolution maps of material stocks in buildings and infrastructures in Austria and Germany. *Environmental Science and Technology*, *55*(5), 3368–3379. https://doi.org/10.1021/acs.est.0c05642

Høyer, K. G., & Holden, E. (2003). Household consumption and ecological footprints in Norway—Does urban form matter? *Journal of Consumer Policy*, *26*(3), 327–349. https://doi.org/10.1023/A:1025680422704

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (3rd ed.). OTexts.

Jiang, L., & O'Neill, B. C. (2017). Global urbanization projections for the shared socioeconomic pathways. *Global Environmental Change*, *42*, 193–199. https://doi.org/10.1016/j.gloenvcha.2015.03.008

Keßler, C. (2015). OpenStreetMap. *Encyclopedia of GIS* (2nd ed.). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-319-23519-6_1654-1

Koppel, K., Zalite, K., Voormansik, K., & Jagdhuber, T. (2017). Sensitivity of Sentinel-1 backscatter to characteristics of buildings. *International Journal of Remote Sensing*, *38*(22), 6298–6318. https://doi.org/10.1080/01431161.2017.1353160

Li, W., He, C., Fang, J., Zheng, J., Fu, H., & Yu, L. (2019). Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. *Remote Sensing*, *11*(4), 403. https://doi.org/10.3390/RS11040403

Li, W., Goodchild, M. F., & Church, R. (2013). An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. *International Journal of Geographical Information Science*, *27*(6), 1227–1250. https://doi.org/10.1080/13658816.2012.752093

Liu, G., Shih, K. J., Wang, T.-C., Reda, F. A., Sapra, K., Yu, Z., Tao, A., & Catanzaro, B. (2018). *Partial convolution based padding*. https://arxiv.org/abs/1811.11718v1

Meta. (2022). *Facebook data for good high resolution population density maps*. https://dataforgood.facebook.com/dfg/tools/high-resolution-population-density-maps

Microsoft. (2021). *Open dataset of machine extracted buildings in Uganda and Tanzania*. https://github.com/microsoft/Uganda-Tanzania-Building-Footprints

Mohandoss, T., Kulkarni, A., Northrup, D., Mwebaze, E., & Alemohammad, H. (2020). *Generating synthetic multispectral satellite imagery from Sentinel-2*. https://arxiv.org/abs/2012.03108v1

Moran, D., Kanemoto, K., Jiborn, M., Wood, R., Többen, J., & Seto, K. C. (2018). Carbon footprints of 13000 cities. *Environmental Research Letters*, *13*(6), 064041. https://doi.org/10.1088/1748-9326/AAC72A

NIRAS. (2021, December 15). *Increasing the resilience of the North coast and Nile River Delta in Egypt*. https://www.niras.com/projects/increasing-the-resilience-of-the-north-coast-and-nile-river-delta-in-egypt/

Owusu, M., Kuffer, M., Belgiu, M., Grippa, T., Lennert, M., Georganos, S., & Vanhuysse, S. (2021). Towards user-driven earth observation-based slum mapping. *Computers, Environment and Urban Systems*, *89*, 101681. https://doi.org/10.1016/j.compenvurbsys.2021.101681

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Schug, F., Frantz, D., van der Linden, S., & Hostert, P. (2021). Gridded population mapping for Germany based on building density, height and type from earth observation data using census disaggregation and bottom-up estimates. *PLoS One*, *16*(3 March), e0249044. https://doi.org/10.1371/journal.pone.0249044

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., & Quinn, J. (2021). *Continental-scale building detection from high resolution satellite imagery*. ArXiv. https://arxiv.org/abs/2107.12283v2

Smith, S. L., Kindermans, P. J., Ying, C., & Le, Q. V. (2018, November 1). Don't decay the learning rate, increase the batch size. ArXiv. https://arxiv.org/abs/1711.00489v2

Statistical Service Ghana. (2021). *Ghana 2021 population and housing census*. https://census2021.statsghana.gov.gh/gssmain/fileUpload/reportthemelist/2021PHCGeneralReportVol3A_PopulationofRegionsandDistricts_181121.pdf

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One*, *10*(2), e0107042. https://doi.org/10.1371/journal.pone.0107042

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, San Francisco, CA (pp. 4278–4284). AAAI. https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806

Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, *4*(1), 1–4. https://doi.org/10.1038/sdata.2017.4

Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E. B., & Dang, H.-A. H. (2017). *Mapping the world population one building at a time*. World Bank. https://doi.org/10.1596/33700

United Nations Population Fund. (2021). *Egypt population projections*. https://www.unfpa.org/data/EG

Xie, Q., Luong, M. T., Hovy, E., & Le, Q. V. (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). Seattle, WI: IEEE Computer Society. https://doi.org/10.1109/CVPR42600.2020.01070

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fibæk, C. S., Keßler, C., Arsanjani, J. J. & Trillo, M. L. (2022). A deep learning method for creating globally applicable population estimates from sentinel data. *Transactions in GIS*, *00*, 1–29. https://doi.org/10.1111/tgis.12971