Mapping Urban Population Growth from Sentinel-2 MSI and Census Data Using Deep Learning: A Case Study in Kigali, Rwanda

Sebastian Hafner
KTH Royal Inst. of Technology
Stockholm, Sweden
shafner@kth.se

Stefanos Georganos
KTH Royal Inst. of Technology
Stockholm, Sweden
stegeo@kth.se

Theodomir Mugiraneza
University of Rwanda
Kigali, Rwanda
thmugiraneza@gmail.com

Yifang Ban
KTH Royal Inst. of Technology
Stockholm, Sweden
yifang@kth.se

Abstract—To better understand current trends of urban population growth in Sub-Saharan Africa, high-quality spatiotemporal population estimates are necessary. While the joint use of remote sensing and deep learning has achieved promising results for population distribution estimation, most of the current work focuses on fine-scale spatial predictions derived from single date census, thereby neglecting temporal analyses. In this work, we focus on evaluating how deep learning change detection techniques can unravel temporal population dynamics at short intervals. Since Post-Classification Comparison (PCC) methods for change detection are known to propagate the error of the individual maps, we propose an end-to-end population growth mapping method. Specifically, a ResNet encoder, pretrained on a population mapping task with Sentinel-2 MSI data, was incorporated into a Siamese network. The Siamese network was trained at the census level to accurately predict population change. The effectiveness of the proposed method is demonstrated in Kigali, Rwanda, for the time period 2016–2020, using bi-temporal Sentinel-2 data. Compared to PCC, the Siamese network greatly reduced errors in population change predictions at the census level. These results show promise for future remote sensing-based population growth mapping endeavors. Code is available on GitHub[1].

Index Terms—Population mapping, Sub-Saharan Africa, Siamese network

I. INTRODUCTION

The projections in the World Population Prospects 2022 report suggest that the global population could reach 9.7 billion in 2050 [1]. At the forefront of the anticipated population growth are countries of Sub-Saharan Africa. In light of this, frequent updates of existing population data in that region are crucial, particularly considering that knowledge of population distribution is a necessary requisite for a wide range of applications. For example, population distribution maps provide vital information for vaccination campaigns, disaster response deployment, and urban mobility and transport planning.

In recent years, census-independent (i.e., bottom-up) population mapping using deep learning and satellite imagery has shown promise in providing accurate population estimates. For example, Doupe *et al.* [2] mapped population density at 8 km spatial resolution in Tanzania and Kenya using a Convolutional Neural Network (CNN) based on the VGG architecture and Landsat 7 imagery. Landsat 7 imagery and the VGG-net were also used by Robinson *et al.* [3] to predict population counts in the United States at 1 km spatial resolution. Authors in [4] proposed to fuse Landsat 8 optical data with Sentinel-1 radar data to predict population density at 4.5 km spatial resolution for rural villages in India and demonstrated that dual-branch fusion networks outperform uni-modal networks. Sentinel-2 (S2) MultiSpectral Instrument (MSI) imagery was used by Huang *et al.* [5] to map population distribution at 1 km spatial resolution for the Atlanta, Georgia, and Dallas, Texas metropolitan areas in the United States of America. Recently, Neal *et al.* [6] used WorldView-2 imagery for estimating population in two districts of Mozambique using representation learning. A ResNet was also used in [7] to map population in Sub-Saharan African cities with multisource satellite imagery from Pleiades and S2. Building footprints were further used to improve the geographical transferability of models.

While deep learning-based population mapping from satellite imagery has gained traction in recent years [2]–[7], little attention has been paid to population growth mapping with the exception of [8]. Using a ResNet and Landsat 5 imagery, Zhuang *et al.* [8] performed population growth analysis in China for the 1985-2010 period by mapping population distribution at 1 km spatial resolution with a 5-year interval. However, analyzing population growth by comparative analysis of independently produced population maps, i.e., change detection by Post-Classification Comparison (PCC), is well-known to suffer from the error propagation of the individual population maps. To that end, we propose an end-to-end population growth mapping method to overcome the error propagation of PCC in uni-temporal population maps. This study is, up to the best of our knowledge, the first to map population growth in an end-to-end fashion from satellite imagery.
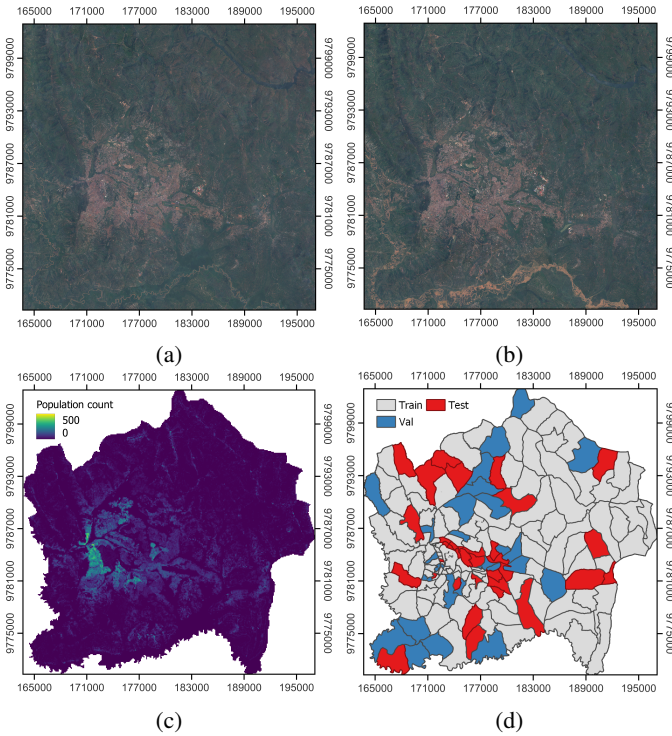
Fig. 1: S2 MSI composites for (a) 2016 and (b) 2020, and (c) population labels at grid level. (d) shows the data set splits.

II. STUDY AREA AND DATA

Kigali, the capital city and economic hub of Rwanda, was selected as the study area. Kigali encompasses an area of approximately 730 km^2. In 2012, Kigali had a population of approximately 1.1 million and placed among the fastest-growing cities in Africa [9]. In recent years, rapid urbanization resulted in the conversion of major cropland areas into built-up areas in the urban fringe zones of Kigali, which increased ecosystem service demands and negatively affected the habitat for biodiversity service function [10].

S2 MSI imagery of Kigali for 2016 and 2020 was retrieved from Google Earth Engine [11]. Specifically, cloud-free composites were generated by collecting all S2 Level-1C (top-of-atmosphere) scenes acquired during the wet season of the respective year. Thereafter, cloudy pixels (i.e., cloud probability > 50 %) were masked for each scene, before the scenes were combined using median compositing. The resulting cloud-free composites for 2016 and 2020 are visualized in Figure 1a and Figure 1b, respectively.

Population census data at the level of designated census enumeration areas were acquired for Kigali for the years 2016 and 2020 (161 administrative polygons). These areas are corresponding to the smallest administrative entities in Rwanda called villages. The data consist of number of population (head counts) and were acquired from two institutions including Kigali city One Stop center and the Local Administrative Entities Development Agency. Using iterative merging, we

aggregated the dataset into a smaller number of units, to reflect a more realistic scenario regarding data availability, but also to adapt to the needs of the experiment (i.e., 100 meter predictive spatial resolution). Finally, the census units were randomly split into a training, validation, and test set (60/20/20 split) (Figure 1d).

III. METHODOLOGY

A. Problem Setup

We consider two S2 MSI images that cover the same geographical area (Kigali) but were acquired at two different times, t_1 and t_2. Furthermore, we consider the census units constituting the City of Kigali, where each census unit, U, contains an accurate count of the population, Y, for t_1 and t_2. The goal is to train a network that accurately predicts the population growth for a census unit $D (= Y^{t_2} - Y^{t_1})$ from the part of the S2 images I^{t_1} and I^{t_2} covering U. However, each census unit has a unique non-rectangular shape and, therefore, cannot be used directly as network input. A common way to deal with this is to operate on a grid level by dividing the entire study area into patches that constitute the census areas [7]. Consequently, census units are composed of a varying number of patches (100 x 100 m). Therefore, the network input to predict the population growth for a census unit is, in practice, the collection of S2 patches, x^{t1} and x^{t2}, constituting the census unit.

B. Proposed Method

The proposed population growth mapping method consists of two stages: 1) an encoder model is pretrained by mapping population at the grid level, and 2) a Siamese network, incorporating the pretrained encoder, is trained at the census level to map population growth.

Population Mapping at Grid Level: Our previous work demonstrated that an encoder based on the ResNet-18 architecture suffices to learn salient features from S2 MSI imagery for population mapping [7]. The same architecture is employed in this work (Figure 2). Specifically, the first layer of the ResNet-18 encoder is replaced with a 3 x 3 conv layer with 4 input channels to accommodate the 10 m S2 bands (Band 2, 3, 4, and 8) as input, while the remaining conv blocks constituting the encoder remain unchanged. The features extracted with the encoder are converted to a population prediction, p, using a fully connected layer. Finally, the ReLu activation function is used to constrain p values to positive numbers.

Hyper-parameters for training are tuned on the validation set using grid search with 3 learning rates (10^{-5}, 10^{-4}, and 10^{-3}) and 2 batch sizes (8, 16). AdamW is used as optimizer, and the training duration is set to 100 epochs with early stopping (patience 5) to prevent models from overfitting to the training set. As in [7], flips (horizontal and vertical) and rotations ($k *$ 90°, where $k \in \{0, 1, 2, 3\}$) are applied to the training data for data augmentation, and the Mean Square Error (MSE) loss (commonly known as L2 loss) is used as loss function. L2 loss is defined as follows: $L2 = (y - p)^2$, where the true and predicted population value is denoted by y and p, respectively.

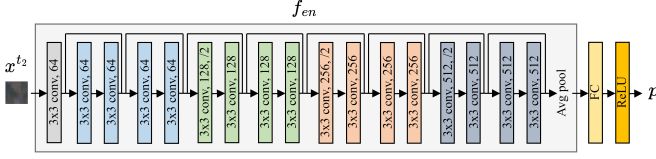An NVIDIA GeForce RTX 3090 graphics card is used for training.

Fig. 2: Diagram of the ResNet-18 model used for grid-level population mapping.

Population Growth Mapping at Census Level: For population growth mapping, we incorporate the pretrained ResNet-18 encoder into a Siamese network (Figure 3). Siamese networks consist of two encoders with shared weights that are used to separately extract features from the inputs, before deriving the change information from the combined features. Due to their inherent suitability to detect differences, Siamese networks have also become a popular architecture for change detection in bi-temporal pairs of satellite images. In this work, the pretrained encoder is employed to extract features on population count from both images separately. The pair of bi-temporal features is then converted to a population growth prediction using a fully connected layer. No activation function is applied to the output of that layer to allow for negative growth predictions.

An important challenge of supervised population growth mapping is that bi-temporal population counts are required for the derivation of growth labels. While it is possible to accurately disaggregate a census to a grid, this requires auxiliary data such as land cover maps or building footprints. However, this data is often not available for both timestamps. Therefore, the Siamese network is trained at the census level by adapting the weakly supervised learning strategy proposed in [12]. Specifically, Metzger *et al.* [12] trained a population mapping model using population count at the census level as labels by comparing them to the aggregated model predictions (patch-level) for corresponding census units. Likewise, we use the Siamese network to predict population growth separately for all patches of a census unit, before applying the loss to the sum of predicted growth, D, using ΔY as label. The training setup (i.e., hyper-parameter tuning, early stopping, and data augmentations) is identical to that for population mapping. It should be noted, however, that the pretrained encoder is frozen during training, meaning that only the fully connected layer (f_{fc} in Figure 3) is trained.

C. Accuracy Metrics

We make use of three commonly employed metrics in population studies [13], namely the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), and the coefficient of determination (R^2). RMSE and MAE are defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - p_i)^2}{n}}, \ \text{MAE} = \sqrt{\frac{\sum_{i=1}^{n}|y_i - p_i|}{n}}, \tag{1}$$
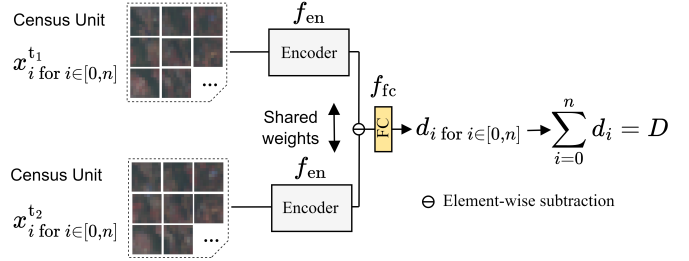

Fig. 3: Diagram of the proposed population growth mapping method consisting of two pretrained ResNet-18 encoders, f_{en}, with shared weights and a fully connected layer, f_{fc}. The network is trained at the census level with frozen encoders.

where y and p are true and predicted values, respectively, and n is the sample size. On the other hand, R^2 is defined as 1 minus the fraction of the residual sum of squares and the total variability of the data.

IV. RESULTS

Table I lists the quantitative population mapping results at the grid level for 2020 and at the census level for 2016 and 2020. All three accuracy metrics indicate that accurate population predictions were achieved at the grid level. However, the aggregated results at the census level provide a stronger validation since the census population counts are official data. While RMSE and MAE values are not comparable between the grid and census level, the R^2 values at the census level indicate good performance (0.70 +), although worse than the performance achieved at the grid level (0.84). It is also apparent that the obtained accuracy values for 2016 and 2020 are relatively similar. Consequently, applying the model to new data from a different year had little impact on model performance.

TABLE I: Quantitative population mapping results at the gird and census level for the test set.

Level	RMSE ↓		MAE ↓		R^2 ↑	
	2016	2020	2016	2020	2016	2020
Grid	-	19	-	10	-	0.84
Census	3,199	3,253	2,368	2,196	0.72	0.73

Figure 4 quantitatively compares the population growth predictions of (a) the PCC with (b) the proposed end-to-end method. The former, PCC, performed poorly, resulting in very high errors (RMSE = 1,471 and MAE = 1,082). In contrast, the proposed method achieved satisfactory results with an RMSE of 202 and an MAE of 165. In terms of R^2 values, the results are more similar, but better performance was also achieved by the proposed method (0.55 vs. 0.67). However, it is also apparent that the proposed method generally underestimates population growth.

The qualitative population growth mapping predictions of the proposed method are visualized in Figure 5b, next to the ground truth in Figure 5a. Although the magnitude of growth
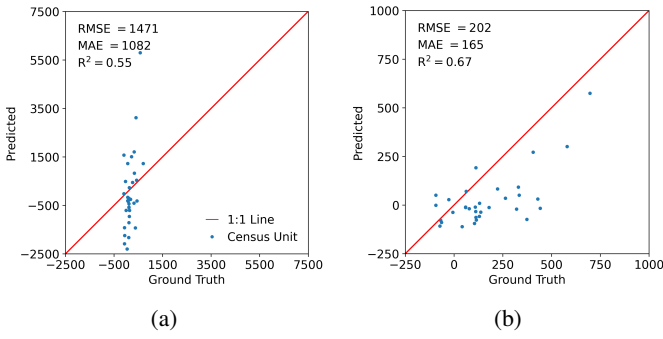
Fig. 4: Population growth test results at the census level for (a) PCC and (b) the proposed method.
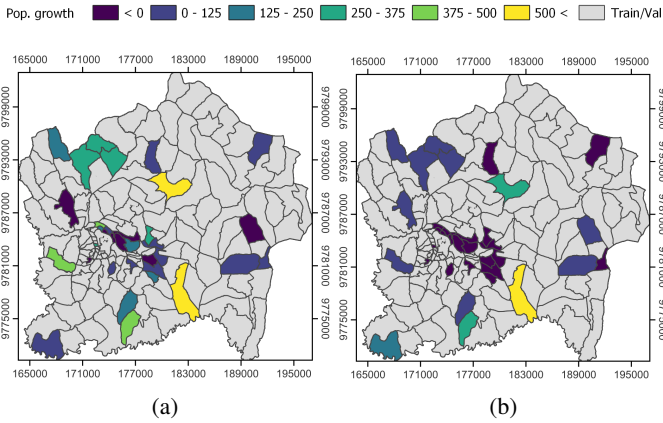

Fig. 5: Population growth maps for the test census units. (a) shows the ground truth and (b) our predictions.

was underestimated, the proposed method picked up on the population growth that occurred on the outskirts of Kigali (e.g., in the northeast and in the central south). However, the model failed to detect population growth in the small census units of central Kigali for which it predicted slightly negative growth values.

V. DISCUSSION AND LIMITATIONS

We find the proposed method to be effective for population growth mapping from S2 MSI imagery, especially compared to PCC. Our findings also emphasize that salient features about population count can be learned from S2 imagery using a ResNet model. These results are in line with [7].

Our work is also subject to several limitations. First of all, to train the Siamese network, bi-temporal census data is required. However, census data, let alone bi-temporal census data, is difficult to obtain in Sub-Saharan Africa, or often not available at all [13]. Moreover, the S2 mission was launched less than 8 years ago, while censuses are typically conducted every 10 years. Consequently, bi-temporal census data for time periods starting after 2015 are largely unavailable. Another limitation of this work is that population predictions are based on the presence of built-up areas, but the land use of these areas may not be residential [3]. To overcome this, Neal *et al.* [6]

suggest including additional data modalities like, for example, night-time light data. Our quantitative results in central Kigali (Figure 5b) also suggest that densification of urban areas, and the subsequent increase in population, may be challenging to accurately predict. Finally, further work is needed to assess if the proposed method can accurately detect negative population growth as a result of, for example, slum evictions.

VI. CONCLUSION

In this paper, a population growth mapping method based on a Siamese network is proposed and evaluated in Kigali, Rwanda for the time period 2016–2020. Using S2 MSI data as input, the proposed method achieved satisfactory population growth mapping results at the census level (RMSE = 202, MAE = 165, R^2 = 0.67), and greatly outperformed PCC in terms of RMSE (-1,269) and MAE (-917).

Our future work will extend the study area to other Sub-Saharan African cities. Furthermore, we will investigate semi-supervised learning for Siamese network training (e.g., [14]) to reduce the dependence on bi-temporal census data.

REFERENCES

[1] United Nations Department of Economic and Social Affairs, Population Division, "World population prospects 2022: Summary of results," Tech. Rep. UN DESA/POP/2022/TR/NO. 3, 2022.

[2] Doupe, P. et al, "Equitable development through deep learning: The case of sub-national population density estimation," in *Proceedings of the 7th Annual Symposium on Computing for Development*, 2016, pp. 1–10.

[3] Robinson, C. et al, "A deep learning approach for population estimation from satellite imagery," in *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*, 2017, pp. 47–54.

[4] Hu, W. et al, "Mapping missing population in rural india: A deep learning approach with satellite imagery," in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 2019, pp. 353–359.

[5] Huang, X. et al, "Sensing population distribution from satellite imagery via deep learning: Model selection, neighboring effects, and systematic biases," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 5137–5151, 2021.

[6] Neal, I. et al, "Census-independent population estimation using representation learning," *Scientific Reports*, vol. 12, no. 1, pp. 1–12, 2022.

[7] Georganos, S. et al, "A census from heaven: Unraveling the potential of deep learning and earth observation for intra-urban population mapping in data scarce environments," *International Journal of Applied Earth Observation and Geoinformation*, vol. 114, pp. 103013, 2022.

[8] Zhuang, H. et al, "Mapping multi-temporal population distribution in china from 1985 to 2010 using landsat images via deep learning," *Remote Sensing*, vol. 13, no. 17, pp. 3533, 2021.

[9] National Institute of Statistics of Rwanda, "The 2012 population and housing census results," Tech. Rep., 2012.

[10] Mugiraneza, T. et al, "Monitoring urbanization and environmental impact in kigali, rwanda using sentinel-2 msi data and ecosystem service bundles," *International Journal of Applied Earth Observation and Geoinformation*, vol. 109, pp. 102775, 2022.

[11] Gorelick, N. et al, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote sensing of Environment*, vol. 202, pp. 18–27, 2017.

[12] Metzger, N. et al, "Fine-grained population mapping from coarse census counts and open geodata," *Scientific Reports*, vol. 12, no. 1, pp. 20085, Nov 2022.

[13] Linard, C. et al, "Population distribution, settlement patterns and accessibility across africa in 2010," *PloS one*, vol. 7, no. 2, pp. e31743, 2012.

[14] Hafner, S. et al, "Urban change detection using a dual-task siamese network and semi-supervised learning," in *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2022, pp. 1071–1074.