Convolutional Expectation Maximization for Population Estimation

Andrea Pomente [(1)] and David Aleandri [(1)]

(1) University of Rome Tor Vergata, Rome, Italy
{pomente.andrea,davidaleandri}@gmail.com

Abstract. There is a fundamental spatial mismatch in the data available for estimate population from satellite imagery. Spectral reflectances are available for each pixel of an image, but ground reference population data are available only for larger zones, therefore satellite imagery has bigger resolution than ground reference images. The general response has been to build models for the average population density of the zones, utilizing spatially aggregated spectral data. This article reports a new approach to solve this problem where per pixel spectral data are used. The already used expectation maximization algorithm (EM) [1] is paired with a convolutional neural network to improve the resolution of a pre-existent population ground truth provided by the GHS POPULATION GRID (LDS) [5]. We start with the satellite imagery by Sentinel-2 mission and, the regression model we have built, upscales the LDS dataset to 10 meters resolution, the same as Sentinel-2 images. As you can see in Table 1, we obtained an AvgRelDelta of 50.05 for Uganda's rural area and 96.31 for the city of Lusaka in Zambia, while our method scored 87.57 in Overall. This results are computed from a test dataset provided for ImageCLEF 2017 Population Estimation Task [8].

Keywords: Convolutional Neural Networks, Information Retrieval, Image Retrieval, Copernicus data, Automatic Population Estimation, FabSpace 2.0, CLEF

1 Introduction

There is a substantial literature on the application of orbital remote sensing to the estimation of human population and population growth. The levels of accuracy achieved in these studies has been limited by several factors [1–3]. First, the relationship between human habitation and the spectral characteristics of land surface and land cover is inherently indirect and inexact. Second, the problem of estimating a quantitative variable like population density across the spatial dimensions of an image is intrinsically more difficult than the more usual qualitative objectives of remote sensing analysis, such as segmentation or classification. And third, while the resolution of remotely sensed imagery is quite adequate for most demographic purposes, ground reference population data for model development and training is generally only available at a much lower resolution, either

for census-related areal units. To date, the response to this problem of spatial incompatibility of data has been to aggregate the finer resolution spectral data up to the geographical level of the available ground reference populations, and then to build regression models for estimating the population density of these larger spatial aggregate units [4]. Aggregated remote sensing predictor variables have included mean reflectances of individual spectral bands, numbers of pixels in various land-use categories, measures of variability and image texture and various band-to-band ratios and other mathematical transformations of the multispectral data. The availability of demographic data defined the spatial level at which modeling took place (either the elements of a regular grid or some form of census district), and the remote sensing data were spatially aggregated. The aggregation took various forms: averages of individual spectral bands or band ratios, measures of variability and texture, counts or proportions of pixels in different land use classes, and so on. By contrast, this paper examines several potential advantages to be gained by modeling at the level of single pixels rather than larger spatial aggregates.

2 Studied areas and data

This work was based on two areas centered on the Lusaka city in Zambia and on a rural area in Uganda. The satellite imagery is provided by Sentinel-2 mission, in particular the 10 meters resolution RGB and Near-infrared bands are used. This choice was made because of the highest resolution respect to other bands. The LDS was recently produced from Landsat imagery (30 meters resolution) collections through automatic analysis of satellite imagery to produce unprecedented fine scale maps quantifying built-up structures in terms of their location and density. The image processing technology exploits structure (texture, morphology, and pattern) as key information. Population estimates were produced and made available for processing by the Center for International Earth Science Information Network (CIESIN)[1]. These estimates consist of country based layers (one for each of 241 countries) of census and administrative polygons containing estimated residential population. The LDS resolution is 250 meters and in order to use in this work the following steps are done:

- The LDS images are re-projected to the same geo-referentiation (UTM 35-South) of Sentinel-2 imagery
- The image is stretched to fit the same dimensions of Sentinel-2 imagery through nearest neighbor interpolation algorithm
- Each pixel has 10x10 meters of resolution, but the value is still related to an area of 250x250 meters. To correct this problem, an initial redistribution operation is applied following this equation:

$$p_{i,j}^{10x10} = \frac{p^{250x250}}{n}; \; i = 1, \ldots 25; \; j = 1, \ldots 25; \; n = 25 \cdot 25 \tag{1}$$

[1] http://www.ciesin.org/

where $p_{i,j}^{10x10}$ is the pixel value in the new resampled image, obtained in the second step, n is the number of 10x10 resolution pixels inside each 250x250 resolution pixel and $p^{250x250}$ is the pixel value in original LDS imagery, with:

$$\sum_{i=1}^{25}\sum_{j=1}^{25} p_{i,j}^{10x10} = p^{250x250} \tag{2}$$

The equation (1) is performed for each pixel in original LDS imagery.

3 Convolutional EM

The method described in this paper assumes that the relation between pixel reflectance and the number of population is non-linear. To make this relation a convolutional neural network (CNN) [6] is defined, it takes as input the pixel values (RGB and near infrared) and returns the population estimate. The CNN is composed of two convolutional layers with kernel size of 2 and 64 feature maps with strides equal to 1 and a ReLU activation function. Each convolutional layer is followed by a batch normalization layer. The final layers are fully connected; the first one has 128 ReLU neurons and the last one has only a neuron with linear activation function.
We can now refine our initial assigned pixel population estimates, by redistributing population within each pixel away from underestimated pixels and towards overestimated pixels while maintaining the known image total. Intuition suggests [1], that the optimal redistribution, which minimizes the sum of squared residues on the regression line and simultaneously holds the sum of the constant p values, is obtained by making all residuals equal, by adjusting the estimated population as follows:

$$p_{i,j,adj} = p_{i,j,pred} + r \tag{3}$$

where $p_{i,j,pred}$ is the previous iteration population prediction make by CNN described above and r is defined by:

$$r = \frac{\sum_{i=1}^{m}\sum_{j=1}^{m}\left(p_{i,j} - p_{i,j,pred}\right)}{m \cdot m} \tag{4}$$

where $m \cdot m$ is the total number of pixels in the input image.
This procedure is iterated many times to obtain a redistribution of pixel's population to minimize a RMS loss function. To train the CNN at each iteration, Adam [9] is used as optimizer algorithm with default parameters. The iteration process is stopped analyzing the values of R^2 coefficient.
The whole process is written in Python using the Keras library [10] with Tensorflow backend. The output of this process is a new image, with the same size of input, representing the new population for each pixel, also the geodata from the geotiff input is reused for the output. Each area of interest is represented by a shape file. To extract the population of each area we used QGIS importing the shape file and the output image from the CNN.

4 Results

The model was tested using the ImageCLEF 2017 Remote Task [8] evaluation system. This pilot task is part of the imageCLEF 2017 Labs [7] and was introduced this year. It aims at investigating the use of non commercial satellite data as a free and quicker process to estimate the population of an area of interest. The following result were obtained:

Team Name	Team Country	Geographic Zone	Sum Delta	RMSE	Pearson	AvgRelDelta
AndreaDavid	Italy	UGD	18,485	1,816	0.76	50.05
AndreaDavid	Italy	ZMB	1,465,603	30,480	0.08	96.31
AndreaDavid	Italy	Overall	1,484,088	2,7462	0.21	**87.57**

Table 1. Population estimation. UGD stands for Uganda while ZMB is for Zambia region. Overall is when considering both regions all together

Looking at the AvgRelDelta metric in the table above, our model performs better for rural area (UGD) than for Lusaka city(ZMB). A possible reason why we have obtained this different can be found in the better initial estimation in LDS ground truth for UGD respect the ZMB area. Considering the short period of development we forced due dead line, we trust the method proposed in this article is promising. It can be surely improved by using better resolution imagery and making more tests on CNN having a test dataset.

Acknowledgments This work is part of FabSpace project that received funding from the European Unions Horizon 2020 Research and Innovation programme under the Grant Agreement n693210. https://www.fabspace.eu/

References

1. Jack T. Harvey: Population Estimation Models Based on Individual TM Pixels. Photogrammetric Engineering & Remote Sensing. Vol. 68, No. 11, November 2002, pp. 1181-1192 (2002)
2. Iisaka, J. and E. Hegedus: Population estimation from Landsat imagery. Remote Sensing of Environment, 12:259-272 (1982)
3. Langford, M., D.J. Maguire, and D.J. Unwin: The areal interpolation problem: Estimating population using remote sensing within a GIs framework. Handling Geographical Information: Methodology and Potential Applications (I. Masser and M. Blakemore, editors), Longman, London, U.K., pp. 55-77 (1991)
4. Webster, C.J.: Population and dwelling unit estimates from space. Third World Planning Review, 18(2):155-176 (1996)
5. European Commission, Joint Research Centre (JRC); Columbia University, Center for International Earth Science Information Network - CIESIN (2015): GHS population grid, derived from GPW4, multitemporal (1975, 1990, 2000, 2015). European Commission, Joint Research Centre (JRC) [Dataset] PID: `http://data.europa.eu/89h/jrc-ghsl-ghs_pop_gpw4_globe_r2015a`

6. Matan, O. and Kiang, R. K. and Stenard, C. E. and Boser, B. and Denker, J. S. and Henderson, D. and Howard, R. E. and Hubbard, W. and Jackel, L. D. and LeCun, Y.: Handwritten character recognition using neural network architectures. Proc. of the 4th US Postal Service Advanced Technology Conference (1990)
7. B. Ionescu, H. Muller, M. Villegas, H. Arenas, G. Boato, D.-T. Dang-Nguyen, Y. Dicente Cid, C. Eickhoff, A. Garcia Seco de Herrera, C. Gurrin, B. Islam, V. Kovalev, V. Liauchuk, J. Mothe, L. Piras, M. Riegler, and I. Schwall. Overview of ImageCLEF 2017: Information extraction from images. In CLEF 2017 Proceedings, volume 10456 of Lecture Notes in Computer Science, Dublin, Ireland, September 11-14 2017. Springer.
8. Arenas, Helbert and Islam, Bayzidul and Mothe, Josiane: Overview of the ImageCLEF 2017 Population Estimation Task. CLEF 2017 Labs Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, September 11-14 2017, Dublin, Ireland.
9. Kingma, D. P., and Ba, J. L. (2015). Adam: a Method for Stochastic Optimization. International Conference on Learning Representations, 113.
10. Chollet François and others: Keras, `https://github.com/fchollet/keras` (2015)