# Estimating Population Using Nighttime Lights Satellite Data

PETER FORTUNA, Brigham Young University, USA

Socioeconomic data is relatively scarce, especially in developing countries. The goal of this project was to predict population using NASA's Black Marble dataset. While the predictions had some success, cloud cover and other interference ultimately prevent the model from predicting useful results. There are many avenues left to explore this approach.

Additional Key Words and Phrases: neural networks, CNN, UNET, geospatial

## 1 INTRODUCTION

Data is crucial to make informed decisions for businesses, NGOs, and state governments. While the amount of data available is rapidly increasing, socioeconomic data remains relatively scarce and messy, especially in developing countries. Despite the high cost of conducting surveys and relatively low amounts of technology in developing countries, there is an increasing amount of satellite data. The goal of this project is to extract meaningful insights from nighttime light satellite data, specifically in regards to population. Several papers how conducted similar research in the context of population, wealth, and natural characteristics like forestation and elevation (Kattenborn, Rolf, Yeh). This project also predicts population, but is unique for the data and model that it uses.

## 2 DATA OVERVIEW

The data-set was created using satellite images from NASA and a population density data set, also published by NASA. These datasets are called VNP46A2 - VIIRS/NPP 500m from NASA's Black Marble nighttime light database, and Gridded Population of the World (GPW, v4) from NASA's Socioeconomic Data and Applications Center (SEDAC). Black Marbel data is updated daily and GPW data is updated yearly. The nighttime light satelite data I used was from Jan 1 2020, and the population data I used was for 2020.

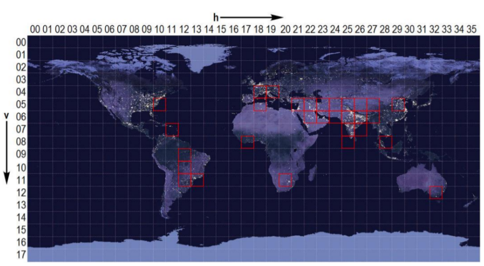


Fig 1. An overview of Black Marble satellite data

Both these datasets use the same latitude and longitudinal scales, so I was able to make a pixel to pixel map where each pixel represents 0.25 square Kilometers. This dataset covers a total of 100 million square kilometers, and used approximately 6.5 Gigabytes of space. Because of this, managing memory was a large focus in my model.

Author's address: Peter Fortuna, peter.j.fortuna@gmail.com, Brigham Young University, Provo, Utah, USA, 84604.

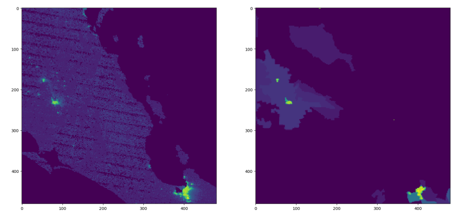

Fig 2. A sample image from my dataset. The left image shows nighttime lights, the right image shows population.

## 3 MODEL

I originally intended to use a modified CNN approach that was presented in "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa". However, this code was difficult to work with and not well suited to the data I found. I eventually decided to modify the UNET cancer lab code instead.

The UNET is an effective and efficient way to implement image detection, or in this case, regression. Instead of using the code to predict the likilihood of cancer or not cancer, I use the UNET to predict the population of each pixel (1/4 square Kilometer).

I used a standard boiled plate training model that involved cycling though several epochs of training images and measuring the loss with test images.

To partition my data, I made every fifth image a test image. As images are sorted geographically, this should help the model to learn general features of each region without learning the exact features of the image it's testing. This should also help the model learn specific patents of static and how to avoid them.

I used ADAM to optimize my model, and MSE as my objective function. MSE punishes outlying predictions relatively highly, and I thought this objective function would create more incentive for the model to learn to predict high population areas. Much of the data contains low and no population areas, so an objective function that does not emphasise high population sections would not create enough incentive to predict any high population regions.

This model did not use any pre-trained weights.

## 4 RESULTS

The model was able to train, but with varied results. It was unable to meaningfully distinguish between noise and city lights in the samples, perhaps due to the noise being inconsistent or the city lights being relatively rare in the images. This could also be due to a dataset that was overall too small, in which case data from more days could be added.

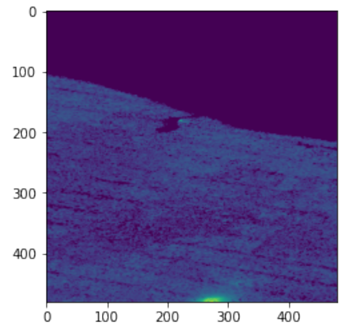

Fig 3. The test image. Note that most of this image is filled with noise.

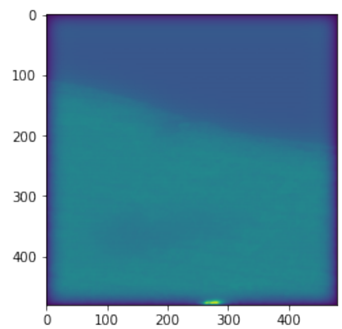

Fig 4. The predicted population. Predictions seem to match the noise in the image more than the actual population.

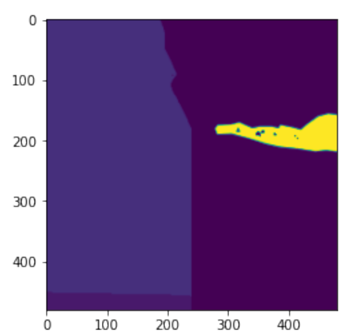

Fig 5. The actual population. The population does not closely match the predicted population.

Notably, the model did not predict better after the first epoch, and seemed to predict any bight area as highly populated, as demonstrated below. This could be due to flaws in the satellite data or flaws in the population label.

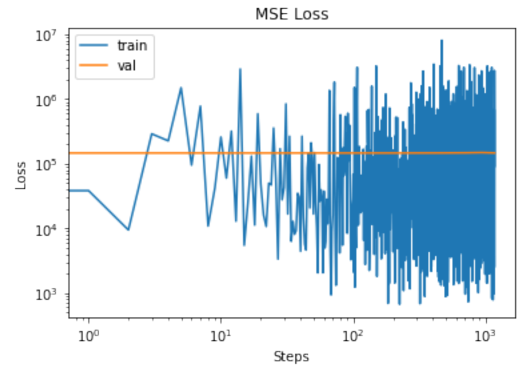


Fig 6. The MSE loss from the train - test cycle. The model's loss improved over time, but only slightly.

## 5 CONCLUSION

While UNETs are powerful tools, I had a hard time training my model to overcome static in the training images.

In the future, I would change my model to predict levels of population. Instead of having continuous output, I think my model would train better and have more interpretable outputs if I did this. I would also be curious to explore techniques for reducing noise in the training data and for training more accurately.

## 6 REFERENCES

Kattenborn, Teja, et al. "Convolutional Neural Networks Enable Efficient, Accurate and Fine-Grained Segmentation of Plant Species and Communities from High-Resolution UAV Imagery." Nature News, Nature Publishing Group, 27 Nov. 2019, https://www.nature.com/articles/s41598-019-53797-9.

Rolf, Esther, et al. "A Generalizable and Accessible Approach to Machine Learning with Global Satellite Imagery." Nature News, Nature Publishing Group, 20 July 2021, https://www.nature.com/articles/s41467-021-24638-z.

Yeh, Christopher, et al. "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa." Nature News, Nature Publishing Group, 22 May 2020, https://www.nature.com/articles/s41467-020-16185-w.