Abstract:
In this work, I develop deep learning computer vision methods that can predict a region's poverty level from an overhead satellite image. I experiment with both daytime and nighttime imagery. Furthermore, because data limitations are often the barrier to entry in poverty prediction from satellite imagery, I explore the impact that data quantity and data augmentation have on the representational power and overall accuracy of the networks. Lastly, to evaluate the robustness of the networks, I evaluate them on data from continents that were absent in the development set.
In this paper, I test the hypothesis that deep learning can leverage satellite imagery to reliably predict the poverty level of a region. I assemble a dataset of 88,386 images from 44,193 cities spanning Africa, South America, Asia, Europe, and the Caribbean. For each city, I obtain a daytime satellite image, a nighttime satellite image, and the city's wealth index. I then train deep neural networks (DNNs) to predict a city's wealth index, given a satellite image. I leverage techniques such as data augmentation, pretraining and transfer learning, regularization, and cross validation in developing and evaluating the networks. I evaluate the networks using cross-country and within-country wealth data, and also explore the impact of data quantity on the representational power of the networks.

Literary review-

- Deep learning has been applied to satellite imagery in the context of structure classification [33], map segmentation [31] [6], and pattern identification [5]. Using satellite imagery to predict poverty levels of cities is not a new task.
- Previous work has explored predicting humanitarian indicators using satellite imagery. [13] uses satellite imagery to predict NOx emissions, and [32] [11] [19] utilize Nightlight and RGB satellite imagery to predict poverty outcomes with promising results.
- Furthermore, deep learning has been used to infer both spatial and temporal differences in local-level economic well-being using multiple sources of satellite imagery [35]. Jean etl. Al. use CNNs to predict poverty from high resolution satellite imagery of terrain10 and Perez et. al. and Xie et. al. utilize transfer learning and multi-modal networks to predict poverty from low resolution satellite imagery [20, 34].
- More recently, Yeh et. al. presents the utility of satellite imagery for research and policy, demonstrating their potential to create a wealth map for the entire world [35]. Interestingly, Uzkent et. al. presented an approach pairing satellite imagery with geolocated Wikipedia articles to accurately predict the wealth index of a region [30]. Lastly, generative models have been used to discern the distribution of satellite images and identify forgery [16].
  To the best of my knowledge, there have been no prior work to this date that explores the impact of data quantity on deep neural networks, or evaluates such networks on data from different continents than they were trained on. Furthermore, to the best of my knowledge, there have been no prior works that explore both nighttime and daytime imagery.

Poverty ground truth labels-
The study uses ground truth data obtained from UN World Bank datasets, comprising 44,193 coordinates representing poverty levels as wealth indices. These indices are normalized to

the [-2, 2] interval, with higher values indicating greater wealth, while zero represents global median wealth. The task is framed as regression. Anonymity protection includes adding 5km jitter in rural and 2 km in urban coordinates.

Neural Network Development-

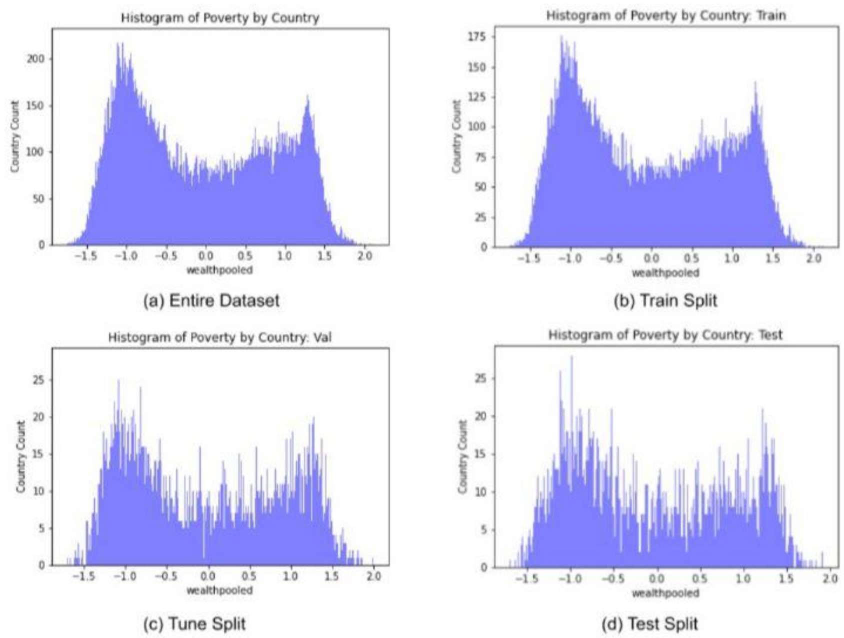
Multiple Deep Neural Networks (DNNs) were trained to predict a region's wealth index, representing poverty. The DNN architecture included flattening the input image and passing it through two multi-layer perceptrons with 512 nodes, yielding a wealth index score. Stochastic Gradient Descent (SGD) with momentum was used, with distinct learning rates for daytime and nighttime images. Training used Mean Squared Error (MSE) loss and was held constant for 10 epochs. PyTorch 1.8 was employed on an NVIDIA Tesla V100 GPU.

Checkpoint selection-

All networks were trained on images only from the train split (See Dataset section). After training, a checkpoint was selected based on what had the smallest RMSE value on the tune set, which consisted of images only from the tune split. Finally, using that checkpoint, inference was run on the test set, which consisted of images only from the test split. Only test set performance is reported.

Evaluation and Statistical analysis-

My main evaluation metric used in the analysis is root mean squared error (RMSE) which is one of the most commonly used metrics to assess performance on regression tasks [7]. The goal is to minimize RMSE values. Confidence intervals (CI) for all evaluation metrics are calculated using the non-parametric bootstrap method with n=1000 permutations at the example level. 95% confidence intervals are reported for all root mean squared errors (RMSE).



Conclusion-

The study demonstrates that Deep Neural Networks (DNNs) can effectively estimate a region's poverty level from satellite images with Root Mean Square Error (RMSE) values slightly above 1. Notably, DNNs perform better on nighttime satellite imagery, which correlates with economic activity. However, daytime images can also provide accurate results with an RMSE as low as 1.477. Increasing data quantity and using data augmentation improve DNN performance, especially for nighttime networks. Generalization experiments show DNNs can generalize well across continents, except when trained on European data and evaluated on African data, indicating significant regional differences. Additional complexity may enhance daytime image performance.