Combining Satellite Imagery and Machine Learning to Predict Poverty

by Neal Jean et al., 2016

Yongjun Zhang

School of Sociology
University of Arizona

Feb 7, 2019

THE UNIVERSITY OF ARIZONA

ECONOMICS

Combining satellite imagery and machine learning to predict poverty

Neal Jean,[1,2*] Marshall Burke,[3,4,5*†] Michael Xie,[1] W. Matthew Davis,[4] David B. Lobell,[3,4] Stefano Ermon[1]

Reliable data on economic livelihoods remain scarce in the developing world, hampering efforts to study these outcomes and to design policies that improve them. Here we demonstrate an accurate, inexpensive, and scalable method for estimating consumption expenditure and asset wealth from high-resolution satellite imagery. Using survey and satellite data from five African countries—Nigeria, Tanzania, Uganda, Malawi, and Rwanda—we show how a convolutional neural network can be trained to identify image features that can explain up to 75% of the variation in local-level economic outcomes. Our method, which requires only publicly available data, could transform efforts to track and target poverty in developing countries. It also demonstrates how powerful machine learning techniques can be applied in a setting with limited training data, suggesting broad potential application across many scientific domains.

Reliable data on economic livelihoods remain scarce in developing world.

- 39 of 59 African countries had fewer than 2 national representative consumption/income-based surveys in 2000-2010.
- 20 of the 59 countries had no DHS asset-based surveys in 2000-2010.

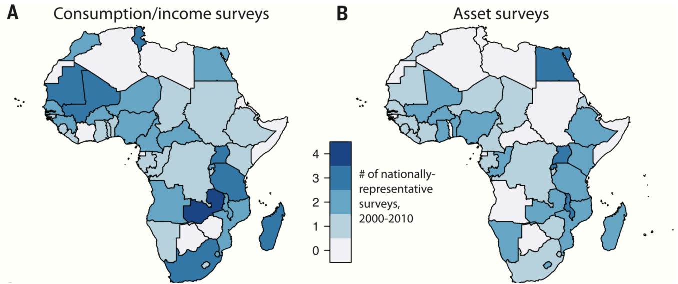


Figure: Poverty data gaps

We need a "data revolution" to sharply scale up data collection efforts within developing countries.

But closing data gaps with more frequent household surveys costs hundreds of billions of U.S. dollars.

Table 2 Population Survey Estimated Costs

Estimated Costs	Small Population	Medium Population	Large Population
Population range:	1 - 5 Million	5 - 20 Million	20+ Million
(WDI Database 2012)			
Census (every 10 years)	$1/ Person	$2/Person	$3/Person
(VSS 2014)			
LSMS (every 5 years)	$0.4 Million	$0.9 Million	$1.5 Million
(Sette 2008; United Nations 2005, 534; Randramamoniy 2008, 1; United Nations 2013c)			
DHS (every 5 years)	$0.8 Million	$1 Million	$1.2 Million
(Yansanch 2000, 771; Rommelmann 2005, 20; WHO 2009, 2)			
CWIQ (annually)	$330,000/ Year	$500,000/ Year	$665,000/ Year
(PARIS 21 2000, 24; Sette 2008)			
MICS (annually)	Financial data not disclosed. No estimates available.		

Source: Morten Jerven, 2014

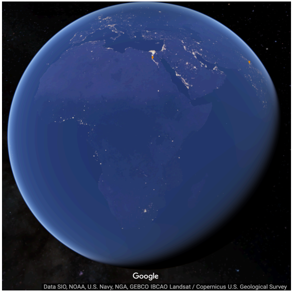
An alternative solution is to use passively collected data.

ECONOMICS

Predicting poverty and wealth from mobile phone metadata

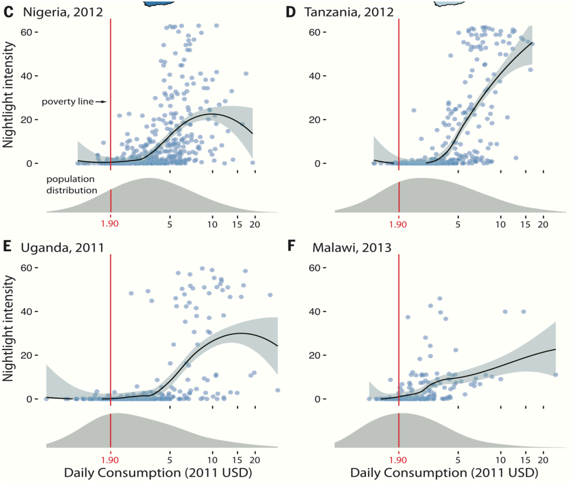
Joshua Blumenstock,[1*] Gabriel Cadamuro,[2] Robert On[3]

Accurate and timely estimates of population characteristics are a critical input to social and economic research and policy. In industrialized economies, novel sources of data are enabling new approaches to demographic profiling, but in developing countries, fewer sources of big data exist. We show that an individual's past history of mobile phone use can be used to infer his or her socioeconomic status. Furthermore, we demonstrate that the predicted attributes of millions of individuals can, in turn, accurately reconstruct the distribution of wealth of an entire nation or to infer the asset distribution of microregions composed of just a few households. In resource-constrained environments where censuses and household surveys are rare, this approach creates an option for gathering localized and timely information at a fraction of the cost of traditional methods.



But phone data are not in public domain.

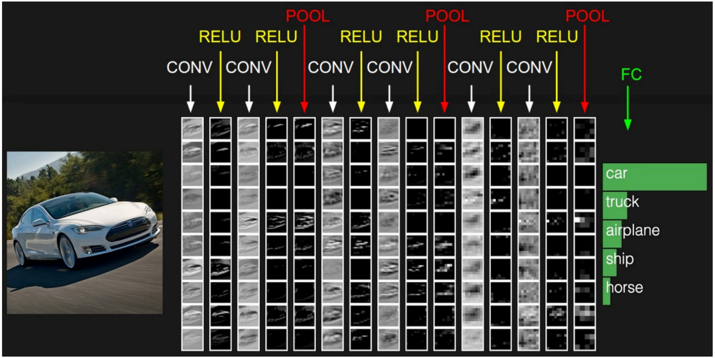
Nightlights often show little variation in poor areas.



Another novel approach is to use high-resolution daytime satellite imagery.

- Satellite imagery is increasingly available at the global scale.
- It contains information on landscape features that may be correlated with economic activities.

The recent development of deep learning methods enables scholars to extract complex features from images.



The activations of an example ConvNet architecture. The initial volume stores the raw image pixels (left) and the last volume stores the class scores (right). Each volume of activations along the processing path is shown as a column. Since it's difficult to visualize 3D volumes, we lay out each volume's slices in rows. The last layer volume holds the scores for each class, but here we only visualize the sorted top 5 scores, and print the labels of each one. The full web-based demo is shown in the header of our website. The architecture shown here is a tiny VGG Net, which we will discuss later.

Source: http://cs231n.stanford.edu/

We can use satellite images to train a deep learning model to estimate economic outcomes of interest.

But we cannot directly train a large convolutional neural network to estimate economic outcomes of interest.

Because we lack enough labeled training data.

- The pretrained CNN model has over 55 million paramters, but we only have several hundred data points for consumption or assets in each country.
- The task of estimating economic well-being from satellite imagery is nontrivial for human non-experts, precluding the generation of additional labeled training data through crowdsourcing services such as Amazon Mechanical Turk.

We develope a multiple step "transfer learning" approach to overcome the scarcity of training data on economic outcomes.

- The basic idea is to train a CNN model on the data-rich nighttime light, a proxy of poverty.
- By solving this related proxy task, the model learns how to extract features that are also useful for the poverty estimation task.

The first step of transfer learning is to obtain the pretrained CNN model from ImageNet that can identify low level image features.

- Obtain an 8-layer CNN model (VGG F) previously trained on the ImageNet.
- The CNN model has over 55 million parameters, and is sufficiently flexible to extract complex features from images.

The second step is to fine-tune the pretrained CNN model to predict the nightlights intensities.

- A classification task with three nighttime lights intensity classes: low (0-2), medium (3-34) and high (35-63).
- Imputs are 400X400 pixel daytime satellite images from Google Static Maps at zoom level 16 (roughly 1 square km areas) in 2013-2015.
- Over 300,000 locations in Africa sampled near DHS locations.

The third step is to use pretrained CNN model to extract satellite image features.

- Use pretrained CNN as a feature extractor for daytime satellite image by discarding the last nighttime light classification layer.
- For each household cluster, using (up to) 100 input images that cover a 10km by 10 km areas centered around the cluster location.
- This results in 100 feature vectors for each cluster.
- We average these feature vectors to obtain one feature vector for the cluster.

Here is a visual example of extracted features.

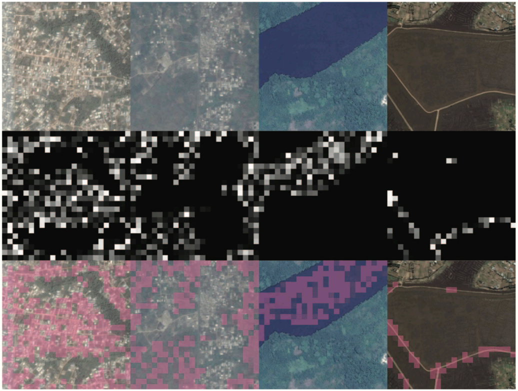


Fig. 2. Visualization of features. By column: Four different convolutional filters (which identify, from left to right, features corresponding to urban areas, nonurban areas, water, and roads) in the convolutional neural network model used for extracting features. Each filter "highlights" the parts of the image that activate it, shown in pink. By row: Original daytime satellite images from Google Static Maps, filter activation maps, and overlay of activation maps onto original images

The final step is to use extracted image features and survey data to train ridge regression models that can estimate assets or consumption expenditure.

- The CNN extracts 4096-dimensional feature vectors from input satellite images.
- Use Principal Component Analysis to reduce the dimension (first 10).
- Train regularized linear regression models to predict cluster-level assets or consumption.

Dimension reduction is necessary because of the lack of enough data points on economic outcomes.

- Average household expenditures are from World Bank's LSMS surveys and average household wealth from DHS surveys at the cluster level.
- We have limited survey data: 1411 clusters for LSMS (2-20 households) and 3034 clusters for DHS (11-45 households) across five African countries.

The transfer learning model is strongly predictive of both consumption and asset across multiple African countries.

Transfer learning models explain 37-55% of the variation in consumption.

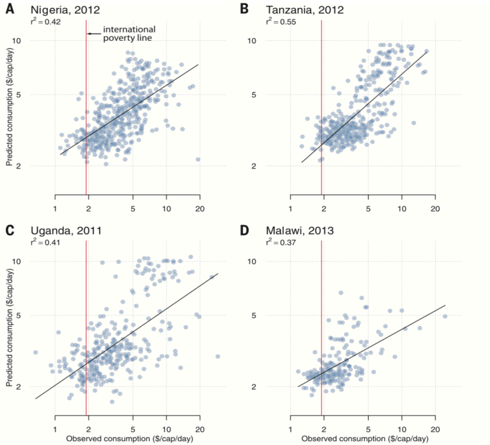
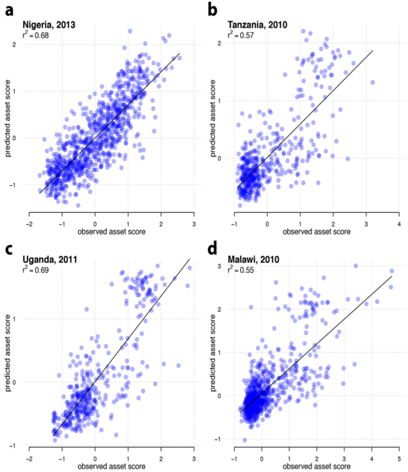


Fig. 3. Predicted cluster-level consumption from transfer learning approach (*y* axis) compared to survey-measured consumption (*x* axis). Results are shown for Nigeria (**A**), Tanzania (**B**), Uganda (**C**), and Malawi (**D**). Predictions and reported r^2 values in each panel are from fivefold cross-validation. Black line is the best fit line, and red line is international poverty line of $1.90 per person per day. Both axes are shown in logarithmic scale. Countries are ordered by population size.

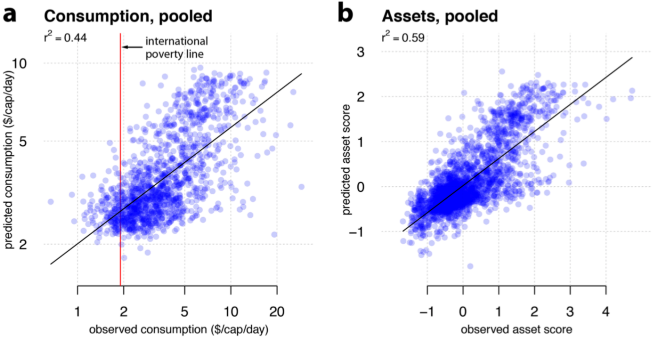
Transfer learning models explain 55-75% of the variation in asset wealth.

Figure S3: **Predicted cluster-level asset index from transfer learning approach (y-axis) compared to DHS-measured asset index (x-axis) for 5 countries.** Predictions and reported r^2 values in each panel are from 5-fold cross validation. Both axes shown in log-scale. Black line is the best fit line.



Pooled Transfer learning models explain 44-59% of the variation in average household consumption and assets across all countries.

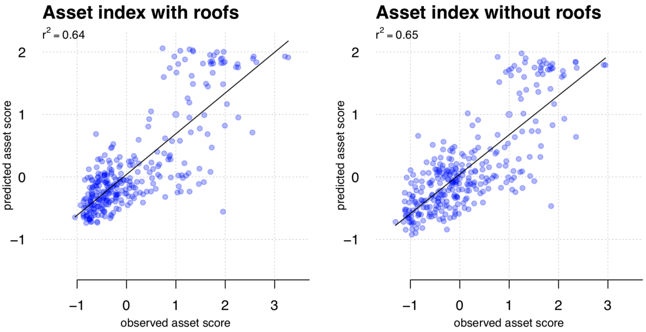
Figure S4: **Relationship between estimated and observed consumption (a) and assets (b)**, from a pooled model using data from all four LSMS countries (as in Figure 3) or all five DHS countries (as in Figure S4). Vertical red line in the left panel is the international poverty line (\$1.90 person^{-1} day^{-1}). Both axes shown in log-scale for consumption.



Predictive power of assets is nearly uniformly higher than for consumption.

- Differences in the outcome being measured, rather than differences in survey design or direct identification of key assets in daytime imagery, likely explain these performance differences.

Figure S5: **Relationship between estimated and observed asset scores, Uganda LSMS**. Left panel uses an asset index that includes variables pertaining to roofing material, right panel omits these variables from asset index. Cross-validated r^2 are reported at top of each panel.



The transfer learning model is on average substantially more predictive of the variation in consumption and assets than nightlights alone.

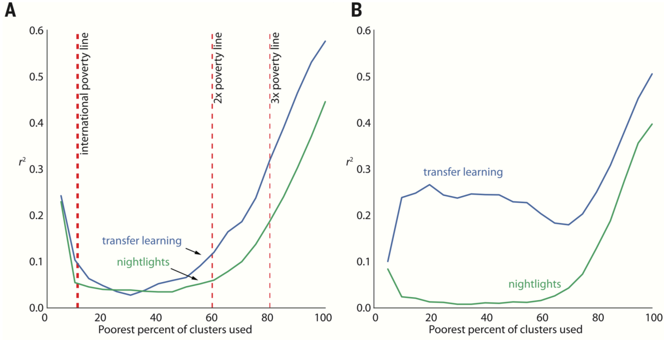
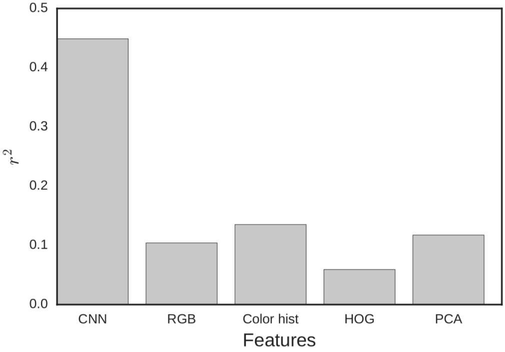


Fig. 4. Evaluation of model performance. (A) Performance of transfer learning model relative to nightlights for estimating consumption, using pooled observations across the four LSMS countries. Trials were run separately for increasing percentages of the available clusters (e.g., x-axis value of 40 indicates that all clusters below 40th percentile in consumption were included). Vertical red lines indicate various multiples of the international poverty line. Image features reduced to 100 dimensions using principal component analysis. (B) Same as (A), but for assets.

The transfer learning model improves upon other simpler approaches to extracting information from daytime imagery and predicting economic outcomes.

Figure S8: **Comparison of CNN and alternative feature extraction methods.** Bar heights represent cross-validated r^2 achieved using five different approaches to feature extraction from daytime satellite imagery. See SM 2.3 for details.



The transfer learning model outperforms using data from past surveys to predict outcomes in recent surveys.

Table S2: **Comparison of our model performance with predictive performance of interpolated earlier DHS surveys.** Second and third columns give the survey year and model performance of our image features. Last column gives the performance of the interpolated predictions based on earlier DHS surveys in the same country, with years of those surveys in parentheses.

Country	**Survey year**	**CNN r^2**	**Interpolated earlier survey r^2**
Uganda	2011	0.69	0.58 (2001), 0.70 (2006)
Tanzania	2010	0.57	0.40 (1999)
Nigeria	2013	0.68	0.43 (1990), 0.50 (2003), 0.70 (2008)
Rwanda	2010	0.75	0.72 (2005)
Malawi	2010	0.55	0.41 (2000), 0.45 (2004)

The transfer learning model's level of predictive performance is unlikely to have arisen by chance.

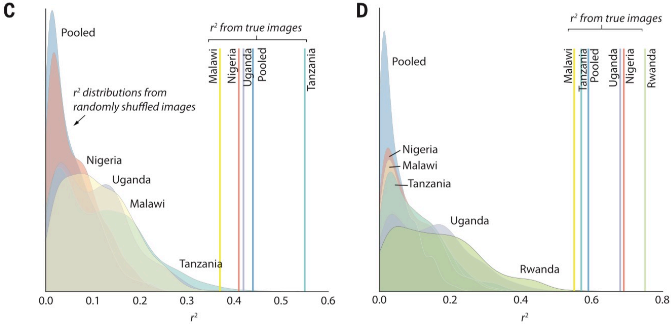


Fig. 4. Evaluation of model performance. (C) Comparison of r2 of models trained on correctly assigned images in each country (vertical lines) to the distribution of r2 values obtained from trials in which the model was trained on randomly shuffled images (1000 trials per country). (D) Same as (C), but for assets. Cross-validated r2 values are reported in all panels.

The transfer learning model travels well across borders, out-of-country predictions often approaching the accuracy of in-country predictions.

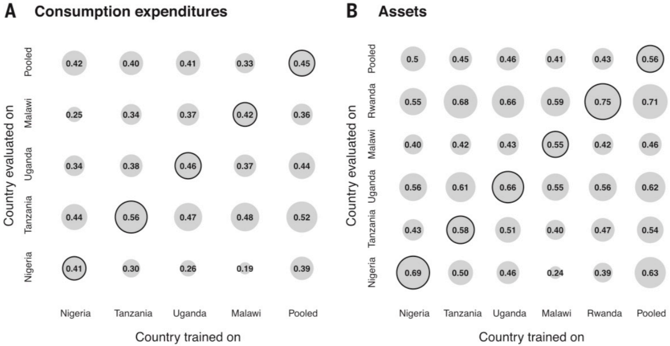


Fig. 5. Cross-border model generalization. (**A**) Cross-validated r^2 values for consumption predictions for models trained in one country and applied in other countries. Countries on x axis indicate where model was trained, countries on y axis where model was evaluated. Reported r^2 values are averaged over 100 folds (10 trials, 10 folds each). (**B**) Same as in (A), but for assets.

High resolution daytime satellite imagery can be used to make fairly accurate predictions about the spatial distribution of economic well-being across five African countries.

The transfer learning approach can be used to fill in the large data gaps resulting from poor survey coverage in many African countries.

THANK YOU!

Sociology ABD at UofA.

Welcome

I am a computational sociologist studying organizational behavior, social networks, and demography. Currently I am a Sociology PhD candidate at the University of Arizona. My dissertation uses statistical, network, and computational methods to study the antecedents and consequences of shareholder activism targeting U.S. corporate political spending practices and policies. My work (will) appears in Journal of Marriage and Family, Demography, Poetics, and Internaltional Journal of Comparative Sociology.

Homepage: https://yongjunzhang.com/