# Temporal Poverty Prediction using Satellite Imagery

**Authors**

Mohammad Mohaiminul Islam (154406)

Mahmudul Hasan (154422)

**Supervisor**

Dr. Md. Hasanul Kabir

Professor

Department of Computer Science and Engineering (CSE)

Islamic University of Technology (IUT)

**A thesis submitted to the Department of CSE
in partial fulfillment of the requirements for the degree of
Bachelor of Science in CSE**

**Department of Computer Science and Engineering (CSE)
Islamic University of Technology (IUT)
Organization of the Islamic Cooperation (OIC)
Gazipur, Bangladesh**

**November 2019**

# Declaration of Authorship

This is to certify that the work presented in this thesis is the outcome of the analysis and experiments carried out by Mahmudul Hasan and Mohammad Mohaiminul Islam under the supervision of Dr. Md. Hasanul Kabir, Professor, Department of Computer Science and Engineering, Islamic University of Technology (IUT), Dhaka, Bangladesh and Rafsanjany Kushol,Lecturer, Department of Computer Science and Engineering, Islamic University of Technology (IUT), Dhaka, Bangladesh. It is also declared that neither of this thesis nor any part of this thesis has been submitted anywhere else for any degree or diploma. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

*Authors:*

| | |
|---|---|
| - - - - - - - - - - - - - - - - - - - - - - - - - | - - - - - - - - - - - - - - - - - - - - - - - - - |
| Mahmudul Hasan | Mohammad Mohaiminul Islam |
| Student ID: 154422 | Student ID: 154406 |

*Supervisor:*

- - - - - - - - - - - - - - - - - - - - - - - - -

Dr. Md. Hasanul Kabir<br>
Professor,<br>
Department of Computer Science and Engineering<br>
Islamic University of Technology (IUT)

# Abstract

Obtaining detailed and reliable data about local economic livelihoods in developing countries is expensive, and data are consequently scarce. Remote sensing data such as high-resolution satellite imagery, on the other hand, is becoming increasingly available and inexpensive. Unfortunately, such data is highly unstructured and currently no techniques exist to automatically extract useful insights to inform policy decisions and help direct humanitarian efforts. Our goal is to extract large-scale socioeconomic indicators from high-resolution satellite imagery.We therefore propose a transfer learning approach where nighttime light intensities are used as a data-rich proxy. We train a fully convolutional CNN model to predict nighttime lights from daytime imagery, simultaneously learning features that are useful for poverty prediction. The model learns filters identifying different terrains and man-made structures, including roads, buildings, and farmlands, without any supervision beyond nighttime lights.

# Acknowledgement

It is an auspicious moment for us to submit our thesis work by which are eventually going to end our Bachelor of Science study. At the very beginning, we want to express our heartfelt gratitude to Almighty Allah for his blessings bestowed upon us which made it possible to complete this thesis research successfully. Without the mercy of Allah, we would not be where we are right now. We would like to express our grateful appreciation to Dr. Md. Hasanul Kabir, Professor, Department of Computer Science and Engineering, Islamic University of Technology and Rafsanjany Kushol,Lecturer, Department of Computer Science and Engineering, Islamic University of Technology (IUT), Dhaka, Bangladesh. for being our adviser and mentor. His motivation, suggestions and insights for this thesis have been invaluable. Without his support and proper guidance, this thesis would not see the path of proper itinerary of the research world. His valuable opinion, time and input provided throughout the thesis work, from the first phase of thesis topics introduction, research area selection, proposition of algorithm, modification and implementation helped us to do our thesis work in proper way. We are grateful to him for his constant and energetic guidance and valuable advice. We would also like to thank Abu Yousuf MD. Abdullah , university of waterloo , Canada, whose work on Land cover identification using satelite imagery inspired us to take on this difficult task. Throughout the research, he extended his helping hands in every way possible. We would like to extend our vote of thanks to all the respected jury members of our thesis committee for their insightful comments and constructive criticism of our research work. Surely they have helped us to improve this research work. Last but not the least, we would like to express our sincere gratitude to all the faculty members of the Computer Science and Engineering department of Islamic University of Technology. They helped make our working environment a pleasant one by providing a helpful set of eyes and ears when problems arose.

# Contents

# List of Figures

# Chapter 1

# Introduction

In this chapter, we first present an overview of our thesis that includes the signification of the problem and the problem statement in detail. Research challenges to be faced in the whole scenario is also discussed based on the problem statement. Thesis objectives, motivations and our contribution are noted in sections. The end of this chapter has the description of the organization of the thesis.

## 1.1 Overview

### 1.1.1 Multi-dimensional poverty index

The Multidimensional Poverty Index (MPI) identifies multiple deprivations at the household and individual level in health, education and standard of living. It uses micro data from household surveys, and—unlike the Inequality-adjusted Human Development Index—all the indicators needed to construct the measure must come from the same survey. Each person in a given household is classified as poor or non-poor depending on the weighted number of deprivations his or her household, and thus, he or she experiences. These data are then aggregated into the national measure of poverty. The MPI reflects both the incidence of multidimensional deprivation (a headcount of those in multidimensional poverty) and its intensity (the average deprivation score experienced by poor people). It can be used to create a comprehensive picture of people living in poverty, and permits comparisons both across countries, regions and the world and within countries by ethnic group, urban or rural location, as well as other key household and community characteristics. The MPI offers a valuable complement to income-based poverty measures. The 2018 Statistical Update presents estimates for 105 developing countries with a combined population of 5.7 billion (77% of the world total). About 1.3 billion people in the countries covered—23.3% of their entire population—lived in multidimensional poverty between 2006 and 2016-17. We could not include other countries due to data constraints. Comparable data on each of the indicators were not available for other developing nations. There was also a decision not to use data from surveys conducted earlier than 2006.

### 1.1.2 Poverty mapping

The Poverty mapping collection enhances our understanding of the geographic distribution of people living in poverty and the conditions of their environment. The data assists policy makers, development agencies, and academic researchers in making decisions to reduce poverty. Spatially explicit data sets are available for selected proxy measures of poverty at global, national, and subnational scales. The global data are of varying resolution, but primarily coarse; the national and subnational data sets are of considerably higher resolution. The data catalogs describe the variables available in each data set, and the underlying spatial, survey, and census data sets used to construct the integrated collection. Details are provided on source data and methods. Poverty maps can be used for several purposes such as highlighting the geographic variations in poverty, understanding the determinants of poverty, designing, targeting, and prioritizing interventions, targeting of programs and projects (infrastructure, health, education, etc.), developing regional policies, prioritization within and between sectors, coordination of sectors, coordination of policies/programs/projects, budget allocation allocation of resources within programs, across projects, allocation of resources across regions, increasing transparency in allocation processes for greater accountability, monitoring and evaluation, provide a baseline on living conditions, monitoring progress in output and outcome indicators, transparency and social accountability, increase transparency in allocation processes, result in greater social accountability.The lack of reliable data in developing countries is a major obstacle to sustainable development, food security, and disaster relief. Poverty data.

### 1.1.3 Land cover

Land cover is the physical material at the surface of the earth. Land covers include grass, asphalt, trees, bare ground, water, etc. Earth cover is the expression used by ecologist Frederick Edward Clements that has its closest modern equivalent being vegetation. The expression continues to be used by the United States Bureau of Land Management. There are two primary methods for capturing information on land cover: field survey and analysis of remotely sensed imagery. Land change models can be built from these types of data to assess future shifts in land cover One of the major land cover issues (as with all natural resource inventories) is that every survey defines similarly named categories in different ways. For instance, there are many definitions of ”forest”—sometimes within the same organisation—that may or may not incorporate a number of different forest features (e.g., stand height, canopy cover, strip width, inclusion of grasses, and rates of growth for timber production).

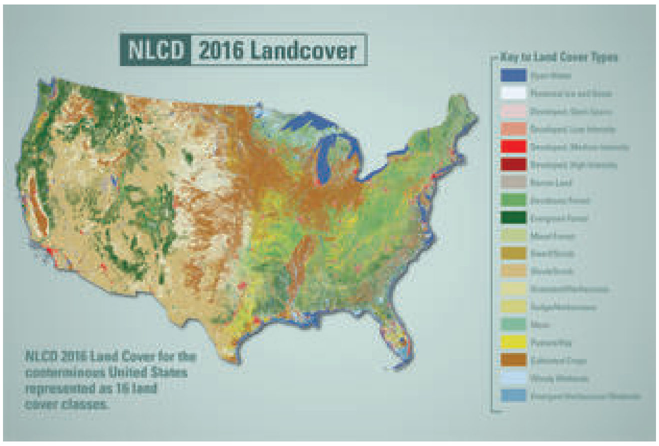


Figure 1: Land cover[4]

### 1.1.4 Landcover classification

Landcover classification system is a comprehensive methodology for description, characterization, classification and comparison of most of land cover features identified anywhere in the world, at any scale or level of detail.LCCS was created in response to a need for:

- a harmonized and standardized collection of land cover data;

- availability of land cover data for a wide range of applications and users;
- comparison and correlation of land cover classes.

## 1.2 Problem Statement

Reliable data on economic livelihoods remain scarce in the developing world hampering efforts to study these outcomes and to design policies that improve them. Obtaining such data is time and labor intensive.We propose a novel machine learning approach to extract large-scale socioeconomic indicators from high resolution satellite imagery.

## 1.3 Research Challenges

1. The main challenge is that training data is very scarce. making it difcult to apply modern machine learning techniques.

2. Dealing with satellite imagery is that the bands of the imagery may have different resolutions.

3. Predicting subnational wealth for new countries, training the models on data for other countries.This task is more difficult, because absolute levels of night light emissions can vary considerably across countries.

4. One of these problems is 'overglow', or the fact that light can spill over from one cell to adjacent cells. This leads to a loss of precision in the night lights data – for example, it is difficult to detect a dark neighborhood in a city due to the spillover of light from bright neighborhoods nearby. We also have lackings in ground truth data .

5. Data only available for last few years and not updated time to time.

6. Spectral complexities of the image objects are introduced due to various unfavorable local environmental and weather conditions (e.g., presence of thick dust, haze, and clouds), and large-scale biophysical changes, resulting from environmental disturbances (e.g., tropical cyclones).

## 1.4 Contributions

The main challenge is that training data is very scarce, making it difficult to apply modern techniques such as Convolutional Neural Networks (CNN). We therefore propose a transfer learning approach where nighttime light intensities are used as a data-rich proxy. We train a fully convolutional CNN model to predict nighttime lights from daytime imagery, simultaneously learning features that are useful for poverty prediction. The model learns filters identifying different terrains and man-made structures, including roads, buildings, and farmlands, without any supervision beyond night time lights.We demonstrate that these learned features are highly informative for poverty mapping, even approaching the predictive performance of survey data collected in the field. By using the data of earlier years we have generated new features such as averaging the feature values from each year and taking the differences of feature values which have improved our accuracy. At first we had to deal with some non-stable light source such as light illumination from forest fire or reflection from the water land for which our model is giving some false positive result , we have corrected this result by merging another model of land cover classification by which we have detected the forest land and the water land.

# Chapter 2

# Literature Review

The first section of this chapter gives a brief overview of the proverty prediction by satelite images with a discussion on some of the precursory works related to remote sensing. In the next section we present different approaches to map proverty.

## 2.1 Background

The lack of reliable data in developing countries is a major obstacle to sustainable development, food security, and disaster relief. Poverty data, for example, is typically scarce, sparse in coverage, and labor-intensive to obtain. Remote sensing data such as high-resolution satellite imagery, on the other hand, is becoming increasingly available and inexpensive. Unfortunately, such data is highly unstructured and currently no techniques exist to automatically extract useful insights to inform policy decisions and help direct humanitarian efforts. So one of the main object is to extract useful insights from the high resolution satelite images.

### 2.1.1 Poverty prediction from night light intensities

One of the approaches was maping proverty level from nightlight intensities [1].In this analysis they find night lights to be good predictors of wealth at the local level.they have categorized the night light intensity into three level low , medium  high. Across the countries they analyzed, the correlation between night light emission and wealth is on average 0.73, and can be as high as 0.87. In order to test the accuracy of their predictions out-of-sample, they set up two prediction tasks. The first is within-country: given a training set with data on wealth and night lights for a number of locations in a country, can they predict the wealth of unseen locations based only on their night light illumination? Predictive performance is very high, indicating that night lights data have great potential for subnational analyses. The second, crossnational task

Figure 2: Night light intensity

is to predict subnational wealth for new countries, training the models on data for other countries. This task is more difficult, because absolute levels of night light emissions can vary considerably across countries . The use of night lights as a proxy for economic variables typically assumes that nighttime illumination corresponds to wealth through one of at least three channels. First, access to the power grid (or a power generator) requires financial investment, which is likely to be made by people with the necessary resources. Second, night lights indicate economic activity, which can lead to higher levels of wealth for the people involved . Third, nighttime illumination (street lamps) can be a result of preferential treatment by the state for certain societal groups Whatever mechanism they assume, high light emissions should be correlated with high levels of wealth. This assumption, however, is not unproblematic. For example, economic activity may not benefit the people living at the location where it occurs – bright commercial centers in cities may be inhabited by poor people. Also, the amount of light emitted by economic activity may not scale directly with the benefits it generates for the local population: oil refineries, for example, are typically illuminated at night, but require few staff and do not coincide with residential areas. the main problem they have faced that they get some false positive value because of non-stable light sources such as forest fire ,reflection from the waterland. they start with a simple example.Figure2 shows an image of the 2008 night lights data for Pakistan(leftpanel).On the right panel, we enlarge the area around Hyderabad and add the corresponding survey data.Intotal,thereare five sampling clusters in the area. The two clusters in the downtown area of the city (two values at the bottom) have bright light values, and at the same time have high wealth scores close to the upper end of the 1–5 scale. As the poor cluster (1.82) shows, dark areas correspond to low wealth values. The topmost point indicates that night lights seemingly capture also intermediate levels of wealth – a moderately bright spot has an intermediate level of wealth (3.75) according to the survey data. Therefore, at least in this initial example, there seems to be a high correlation between night light emissions and wealth at the local level. In order to examine whether this relationship holds more generally, we provide scatter plots for night light emissions (using our default buffer sizes of 2 km for urban and 5 km for rural areas) and wealth for each survey. Figure 2 shows these plots for a selection of three countries (Albania, Cameroon, and Liberia). Complete results are available in Appendix B. The figure provides two important insights. First, the relation between night lights and wealth seems to have a characteristic 'hockey stick' shape, where locations with little to no emissions are generally in the poorest two quintiles of the distribution. The plots in the appendix show that this holds across the vast majority of allcountries/surveys examined in our analysis. Second, while the general shape of the relationship seems to be comparable across countries,the absolute levels of night lights are not. For example, in Albania the richest household clusters emit roughly three times theamount of night lights as those in Liberia. This is something to keep in mind when predicting wealth across countries, as we need to take these differences in magnitude into account.

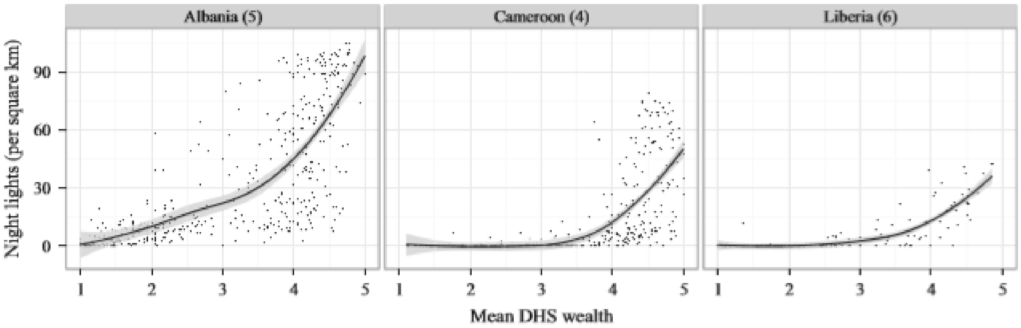


Figure 3: Scatterplot of night lights and wealth for three countries/surveys[2]

### 2.1.2 Poverty prediction through transfer learning

In another approach they have used transfer learning.[4] The convolutional neural network can, in principle, be directly applied to extract socioeconomic factors, but the primary challenge is a lack of training data. they overcome this lack of training data by using a sequence of transfer learning steps and a convolutional neural network model. The idea is to leverage available datasets such as ImageNet to extract features and high-level representations that are useful for the task of interest, i.e., extracting socioeconomic data for poverty mapping. start with a CNN model pre-trained for object classification on ImageNet and learn a modified network that predicts nighttime light intensities from daytime imagery. To address the trade-off between fixed image size and information loss from image scaling, we use a fully convolutional model that takes advantage of the full satellite image. They show that transfer learning is successful in learning features relevant not only for nighttime light prediction but also for poverty mapping. For instance, the model learns filters identifying man-made structures such as roads, urban areas, and fields without any supervision beyond nighttime lights, i.e., without any labeled examples of roads or urban areas. They demonstrate that these features are highly informative for poverty mapping and capable of approaching the predictive performance of survey data collected in the field. CNN models trained on the ImageNet dataset are recognized as good generic feature extractors, with low-level and mid-level features such as edges and corners that are able to generalize to many new tasks. Our goal is to transfer knowledge from the ImageNet object recognition challenge to the target problem of predicting nighttime light intensity from daytime satellite imagery.In the next step they transfer knowledge from the task of predicting nighttime light intensity from daytime satellite imagery to the task of predicting proverty. Here they have used a fully convolutional architecture. In another approach [3] they improved the accuracy by using Landsat 7 Satellite Imagery. They have categoriged night light intesities into 3 levels.Here they gave also a comparison of results using only RGB bands and using all nine bands of Landsat 7 Satellite Imagery.A challenge in dealing with satellite imagery is that the bands of the imagery may have different resolutions, as explained above. A naive workaround is to upsample all bands to the same, highest resolution, which may cause artifacts due to duplicated pixel values and poor utilization of pretrained weights. Instead, we upsample all bands to the highest resolution of 15m/px using Nearest Neighbors and apply dilated convolutions [4] in the first layer of the CNN. At a high level, the goal of modifying the first layer implementation is to preserve the ability to initialize the network from weights pre-trained on RGB image datasets (such as Imagenet) while removing potential artifacts caused by the mismatched resolutions of the multi-spectral imagery. The dilated convolutional layers we implement vary with with the overall model architecture. They also tested several models for predicting poverty metrics from the image features extracted by the CNNs, including ridge regression and gradient-boosted trees (GBTs). We trained each model with leave-one-country-out cross-validation.

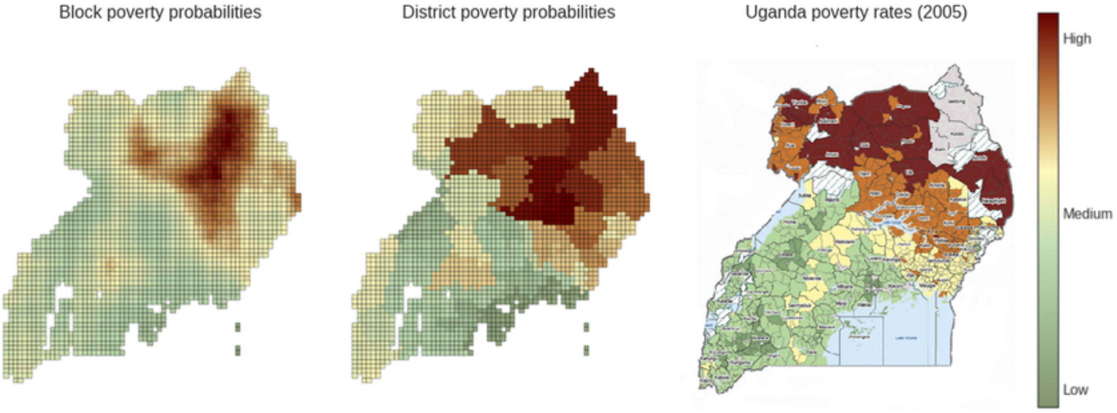

Block poverty probabilities
District poverty probabilities
Uganda poverty rates (2005)
High
Medium
Low

Figure 4: Left: Predicted poverty probabilities at a fine grained 10km * 10km block level. Middle: Predicted poverty probabilities aggregated at the district-level. Right: 2005 survey results for comparison (World Resources Institute 2009).[1]

## 2.2 Methods

### 2.2.1 Convolutional Neural Network

Convolutional Neural Networks are very similar to ordinary Neural Networks from the previous chapter: they are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. The whole network still expresses a single differentiable score function: from the raw image pixels on one end to class scores at the other. And they still have a loss function (e.g. SVM/Softmax) on the last (fully-connected) layer and all the tips/tricks we developed for learning regular Neural Networks still apply. So what changes? ConvNet architectures make the explicit assumption that the inputs are images, which allows us to encode certain properties into the architecture. These then make the forward function more efficient to implement and vastly reduce the amount of parameters in the network.

$$g_i = P_i(f_i(w_i * x + b_i)) \tag{1}$$

where for the i-th convolutional layer,

$$w_i \in R^l * l * \vec{d} \tag{2}$$

Architecture overview : Recall: Regular Neural Nets. As we saw in the previous chapter, Neural Networks receive an input (a single vector), and transform it through a series of hidden layers. Each hidden layer is made up of a set of neurons, where each neuron is fully connected to all neurons in the previous layer, and where neurons in a single layer function completely independently and do not share any connections. The last fully-connected layer is called the “output layer” and in classification settings it represents the class scores. Regular Neural Nets don’t scale well to full images. In CIFAR-10, images are only of size 32x32x3 (32 wide, 32 high, 3 color channels), so a single fully-connected neuron in a first hidden layer of a regular Neural Network would have 32*32*3 = 3072 weights. This amount still seems manageable, but clearly this fully-connected structure does not scale to larger images. For example, an image of more respectable size, e.g. 200x200x3, would lead to neurons that have 200*200*3 = 120,000 weights. Moreover, we would almost certainly want to have several such neurons, so the parameters would add up quickly! Clearly, this full connectivity is wasteful and the huge number of parameters would quickly lead to overfitting. 3D volumes of neurons. Convolutional Neural Networks take advantage of the fact that the input consists of images and they constrain the architecture in a more sensible way. In particular, unlike a regular Neural Network, the layers of a ConvNet have neurons arranged in 3 dimensions: width, height, depth. (Note that the word depth here refers to the third dimension of an activation volume, not to the depth of a full Neural Network, which can refer to the total number of layers in a network.) For example, the input images in CIFAR-10 are an input volume of activations, and the volume has dimensions 32x32x3 (width, height, depth respectively). As we will soon see, the neurons in a layer will only be connected to a small region of the layer before it, instead of all of the neurons in a fully-connected manner. Moreover, the final output layer would for CIFAR-10 have dimensions 1x1x10, because by the end of the ConvNet architecture we will reduce the full image into a single vector of class scores, arranged along the depth dimension. Here is a visualization:

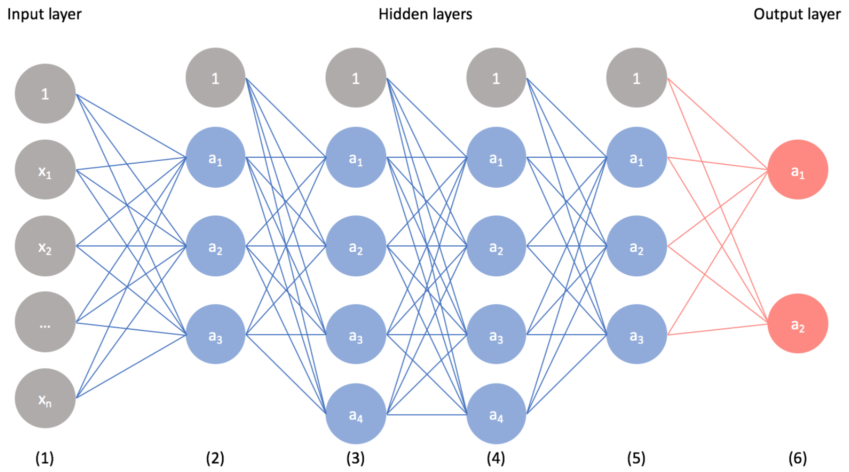


Figure 5: CNN model[42]

### 2.2.2 Transfer learning

Human learners appear to have inherent ways to transfer knowledge between tasks. That is, we recognize and apply relevant knowledge from previous learning experience when we encounter new tasks. We formalize transfer learning as in(PanandYang2010):A domain

$$Domain, D = \{X, P(X)\} \tag{3}$$

Where,
X = feature space
P(X)=Marginal probability distribution
Task equation depending on domain, D

$$T = \{Y, f(.)\} \tag{4}$$

The predictive function

$$f(.) = P(y|x) for\ y \in Y and x \in X \tag{5}$$

The more related a new task is to our previous experience, the more easily we can master it. A tranfer learning graph

$$G = (V, \in); for \tag{6}$$

$$V = \{P_1, ......P_v\}; \tag{7}$$

$$\in = \{(P_i 1, P_j 1), .......(P_i e, P_j e)\} \tag{8}$$

Transfer learning involves the approach in which knowledge learned in one or more source tasks is transferred and used to improve the learning of a related target task. While most machine learning algorithms are designed to address single tasks, the development of algorithms that facilitate transfer learning is a topic of ongoing interest in the machine-learning community.Many deep neural networks trained on natural images exhibit a curious phenomenon in common: on the first layer they learn features similar to Gabor filters and color blobs. Such first-layer features appear not to specific to a particular dataset or task but are general in that they are applicable to many datasets and tasks. As finding these standard features on the first layer seems to occur regardless of the exact cost function and natural image dataset, we call these first-layer features general. For example, in a network with an N-dimensional softmax output layer that has been successfully trained towards a supervised classification objective, each output unit will be specific to a particular class. We thus call the last-layer features specific. In transfer learning we first train a base network on a base dataset and task, and then we repurpose the learned features, or transfer them, to a second target network to be trained on a target dataset and task. This process will tend to work if the features are general, that is, suitable to both base and target tasks, instead of being specific to the base task. In practice, very few people train an entire Convolutional Network from scratch because it is relatively rare to have a dataset of sufficient size. Instead, it is common to pre-train a ConvNet on a very large dataset (e.g. ImageNet, which contains 1.2 million images with 1000 categories), and then use the ConvNet either as an initialization or a fixed feature extractor for the task of interest. Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. It is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skill that they provide on related problems.

# Chapter 3

# Proposed Method

This chapter presents our proposed Temporal Poverty Prediction using Satellite Imagery .The first section provides an overview of the proposed methodology by outlining the components of the system. In the subsequent sections, these components are described in details.

## 3.1 Overview

In the proposed method of Temporal Poverty Prediction from satellite imagery we have used transfer learning and a fully convolutional CNN model. Modern approaches such as Convolutional Neural Networks (CNN) can, in principle, be directly applied to extract socioeconomic factors, but the primary challenge is a lack of training data. While such data is readily available in the United States and other developed nations, it is extremely scarce in Africa where these techniques would be most useful. We overcome this lack of training data by using a sequence of transfer learning steps and a convolutional neural network model . The idea is to leverage available datasets such as ImageNet to extract features and high-level representations that are useful for the task of interest, i.e., extracting socioeconomic data for poverty mapping. Pre-training on ImageNet is useful for learning low-level features such as edges. However, ImageNet consists only of object-centric images, while satellite imagery is captured from an aerial, bird's-eye view. We therefore employ a second learning step, where nighttime light intensities are used as a proxy for economic activity. Specifically, we start with a CNN model pre-trained for object classification on ImageNet and learn a modified network that predicts nighttime light intensities from daytime imagery. We show that transfer learning is successful in learning features relevant not only for nighttime light prediction but also for poverty mapping. For instance, the model learns filters identifying man-made structures such as roads, urban areas, and fields without any supervision beyond nighttime lights, i.e., without any labeled examples of roads or urban areas. We demonstrate that these features are highly informative for poverty mapping and capable of approaching the predictive performance of survey data collected in the field.

Start with a CNN model pre-trained for object classification on ImageNet and learn a modified network that predicts nighttime light intensities from daytime imagery. We show that transfer learning is successful in learning features relevant not only for nighttime light prediction but also for poverty mapping. For instance, the model learns filters identifying man-made structures such as roads, urban areas, and fields without any supervision beyond nighttime lights, i.e., without any labeled examples of roads or urban areas. We demonstrate that these features are highly informative for poverty mapping and capable of approaching the predictive performance of survey data collected in the field.

## 3.2 Framework

The generalized classification steps look like:

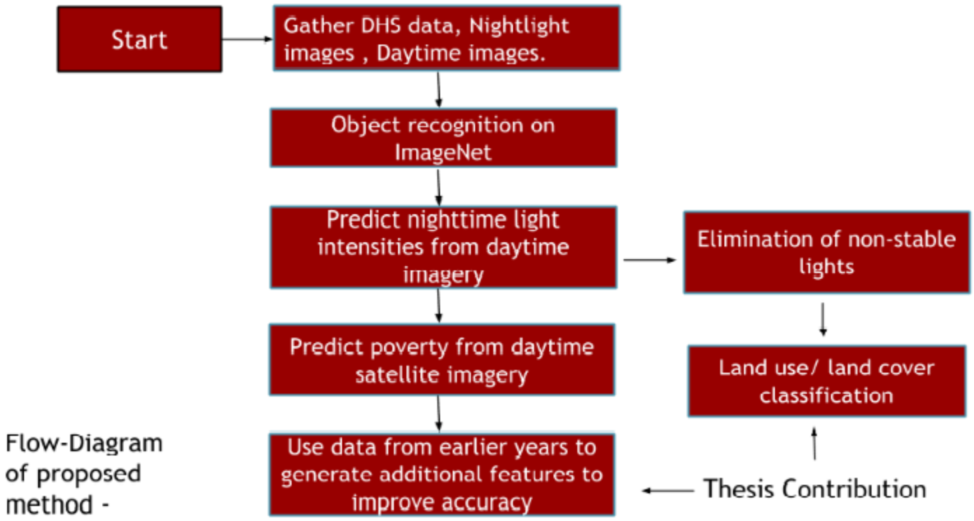

Start
Gather DHS data, Nightlight images , Daytime images.
Object recognition on ImageNet
Predict nighttime light intensities from daytime imagery
Predict poverty from daytime satellite imagery
Use data from earlier years to generate additional features to improve accuracy
Elimination of non-stable lights
Land use/ land cover classification
Thesis Contribution
Flow-Diagram of proposed method -

Figure 6: Temporal poverty prediction framework

## 3.3 Pre-processing

We sample training imagery more densely near locations where labeled survey data is available in order to create more similar image distributions across the transfer learning domains. We take care dividing our sampled images into training, validation, and test splits such that there is no spatial overlap among the splits (though some images within each split may overlap).

## 3.4 Feature extraction

### 3.4.1 Extract features from daytime imagery using deep learning libraries

ImageNet is an object classification image dataset of over 14 million images with 1000 class labels that, along with CNN models, have fueled major breakthroughs in many vision tasks. CNN models trained on the ImageNet dataset are recognized as good generic feature extractors, with low-level and mid-level features such as edges and corners that are able to generalize to many new tasks. Our goal is to transfer knowledge from the ImageNet object recognition challenge to the target problem of predicting nighttime light intensity from daytime satellite imagery. We begin by using a Convolutional Neural Network that has been pre-trained on ImageNet to extract features from the images. We have used the Keras library to use a basic CNN model called VGG16 to extract features of the daytime images. The VGG F model has 8 convolutional and fully connected layers. Like many other ImageNet models, the VGG F model accepts a fixed input image size of 224 * 224 pixels. After feature extraction we have a merged dataset with 492 rows corresponding to the total number of clusters and 12289 columns. Out of the 12289 columns, one column indicates the average cluster wealth and there were a total of 4096 features collected from the daytime images of each year from 2010 to 2012. These were combined into one csv file to form the initial feature set.

### 3.4.2 Use the nightlights to retrain the CNN and extract features

Next, we build on the knowledge gained from this image classification task and fine-tune the CNN on a new task, training it to predict the night time light intensities corresponding to input daytime satellite imagery. Here we use the word “predict” to mean estimation of some property that is not directly observed, rather than its common meaning of inferring something about the future. Nightlights are a noisy but globally consistent—and globally available—proxy for economic activity. In this second step, the model learns to “summarize” the high-dimensional input day time satellite images as a lower-dimensional set of image features that are predictive of the variation in nightlights. Instead of using the image features extracted by the CNN directly to predict wealth, we want to retrain the CNN to predict nightlights from daytime imagery, and use those features, which presumably are more appropriate to our final prediction task i.e. to predict the average wealth. First we divide daytime images into three groups, corresponding to images where the corresponding night-lights pixel is dim, medium, or bright. We define the groups as follows: Dim: nighttime luminosity between 0 and 3. Medium: nighttime luminosity between 3 and 35. Bright: nighttime luminosity between 35 and 64.

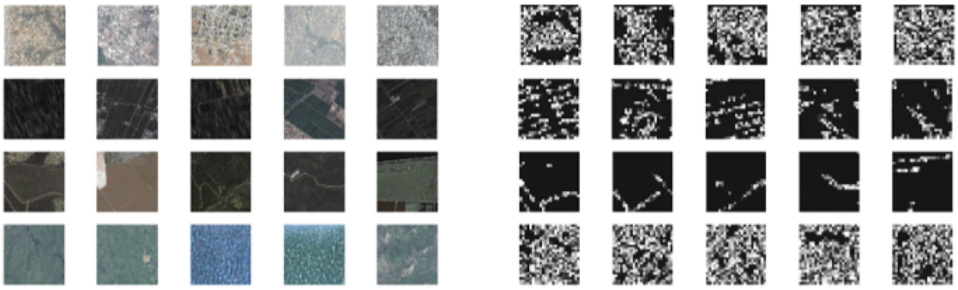


Figure 7: Left: Each row shows five maximally activating images for a different filter in the fifth convolutional layer of the CNN trained on the nighttime light intensity prediction problem. The first filter (first row) activates for urban areas. The second filter activates for farmland and grid-like patterns. The third filter activates for roads. The fourth filter activates for water, plains, and forests, terrains contributing similarly tonight time light intensity. The only super vision used is night time light intensity, i.e., no labeled examples of roads or farmlands are provided. Right: Filter activations for the corresponding images on the left. Filters mostly activate on the relevant portions of the image. For example, in the third row, the strongest activations coincide with the road segments. Images from Google Static Maps.[1]

### 3.4.3 Elimination of non-stable lights

In the process of predicting the average wealth of a cluster , there lies the possibility of having false positives i.e. non – urban areas being classified to be a wealthy area. This may be due certain factors such as forest fires, large camps, reflections from lakes etc. To eliminate this , we devised a two step method of classification and comparison.

**1. Classification:**

First we use land cover classification techniques to classify urban and non-

| Outcome | Score | Meaning |
|---|---|---|
| 1-urban-1-light | 0 | Expected |
| 0-non urban - 0 dark | 0 | Expected |
| 1-urban - 0 dark | 1 | poverty is predicted |
| 0-non urban - 1-light | -1 | False positive |

Table 1: Method of determination of false positive

urban land. For the Daytime images, we label each pixel 1 for Urban and 0 for Non-Urban.The images are segmented using using the scikit-learn quick shift and felzenszwalb modules. We then train the scikit-learn RandomForestClassifier model using training data to classify segments. Thus, we are able to bin the data into urban and non-urban classes.

For the nighttime lights, we label each pixel as 1 for light and 0 for dark. We train the scikit-learn RandomForestClassifier model using training data to classify segments. Thus, we are able to bin the data into light and dark classes.

**2. Comparison:**

• Subtract the Daytime image from the Nighttime image

• Urban areas are expected to have a nighttime imagery falling in 'light' group and non-urban areas are expected to have nighttime imagery falling in 'dark' group. However, if an area is urban but has a corresponding 'dark' nighttime image, then we can be confident that the area has poverty and can also identify false positives for the opposite scenario.

# Chapter 4

# Experimental Analysis

In this chapter we discuss about the experimental setup, the dataset, comparisons and result analysis.

## 4.1 Experimental Setup:

The experiment was conducted in a PC having the following specifications:
- Processor: Intel Core i5 8400
    - o 6 cores
    - o 6 threads
    - o 2.8 GHz (up to 4 GHz)
- RAM: 16GB
- Cache: 9MB L3
- GPU: Aorus GTX 1060

## 4.2 Evaluation Methodologies:

**R-square value**: R-squared (R2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model.

$$R - squared = Explainedvariation/Totalvariation \tag{9}$$

**Accuracy** : Accuracy is the quality or state of being correct or precise. It is the degree to which the result of a measurement, calculation, or specication conforms to the correct value or a standard. In our case, accuracy is the percentage of images whose average wealth was correctly identified.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{10}$$

**F1-score** : F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall. In our case, F1 score is 0.701.

$$F1\ Score = 2 * (Recall * Precision)/(Recall + Precision) \tag{11}$$

So, whenever you build a model, this article should help you to figure out what these parameters mean and how good your model has performed.

**Precision** : Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labeled as survived, how many actually survived? High precision relates to the low false positive rate. We have got 0.788 precision which is pretty good.

$$Precision = TP/TP + FP \tag{12}$$

**Recall :** Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes. The question recall answers is: Of all the passengers that truly survived, how many did we label? We have got recall of 0.631 which is good for this model as it's above 0.5.

$$Recall = TP/TP + FN \tag{13}$$

**AUC :** AUC - ROC curve is a performance measurement for classification problem at various thresholds settings. ROC is a probability curve and AUC represents degree or measure of separability. It tells how much model is capable of distinguishing between classes. Higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s. By analogy, Higher the AUC, better the model is at distinguishing between patients with disease and no disease. The ROC curve is plotted with TPR against the FPR where TPR is on y-axis and FPR is on the x-axis.

$$Area = \int_a^b f(x)dx \tag{14}$$

## 4.3 Dataset:

We have used 3 distinct datasets throughout the experiment to evaluate the effectiveness of our proposed method.

### 4.3.1 Demographic and health surveys data (Average cluster wealth) of the country of Rwanda for 2012.

Demographic and Health Surveys (DHS) are nationally-representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition. There are two main types of DHS Surveys: Standard DHS Surveys have large sample sizes (usually between 5,000 and 30,000 households) and typically are conducted about every 5 years, to allow comparisons over time. Interim DHS Surveys focus on the collection of information on key performance monitoring indicators but may not include data for all impact evaluation measures (such as mortality rates). These surveys are conducted between rounds of DHS surveys and have shorter questionnaires than DHS surveys. Although nationally representative, these surveys generally have smaller samples than DHS surveys.

### 4.3.2 Nighttime luminosity data (Nightlights) of Rwanda of 2012 from National Oceanic and Atmospheric Administration (NOAA).

The National Oceanic and Atmospheric Administration (NOAA) provides annual nighttime images of the world with 30 arc-second resolution, or about 1 square kilometer (NOAA National Geophysical Data Center 2014). The light intensity values are averaged and denoised for each year to ensure that ephemeral light sources do not affect the data.

### 4.3.3 Satellite data from Google Maps API of Rwanda of 2010, 2011, 2012.

Daytime images were downloaded from Google maps at zoom level 16 (pixel resolution is about 2.5m). The image size is set to be 400 pixels X 400 pixels, so that each image downloaded will cover 1 square kilometer. In this way, each daytime image downloaded will correspond to a single pixel from the nighttime imagery.

## 4.4 Result Analysis

### 4.4.1 Test whether night lights data can predict wealth, as observed in DHS

We have "ground truth" measures of average cluster wealth. Our goal is to understand whether the nightlights data can be used to predict wealth. Here we will have 64 luminosity levels. First, merge the DHS and nightlights data, and then fit a ridge regression model of wealth on nightlights. Perform a spatial join to compute the average nighttime luminosity for each of the DHS clusters. To do this, we should take the average of the luminosity values for the nightlights locations surrounding the cluster centroid. Then fit a regression line to illustrate the relationship between cluster average wealth and corresponding cluster nightlights. Now, use cross-validation to get a better sense of out of sample accuracy. We get the R-square value 0.752 from our best model.

### 4.4.2 Test whether basic features of daytime imagery can predict wealth

Earlier we tested whether nightlight imagery could predict the wealth of Rwandan villages. We will now test whether simple metrics (intensity values) from

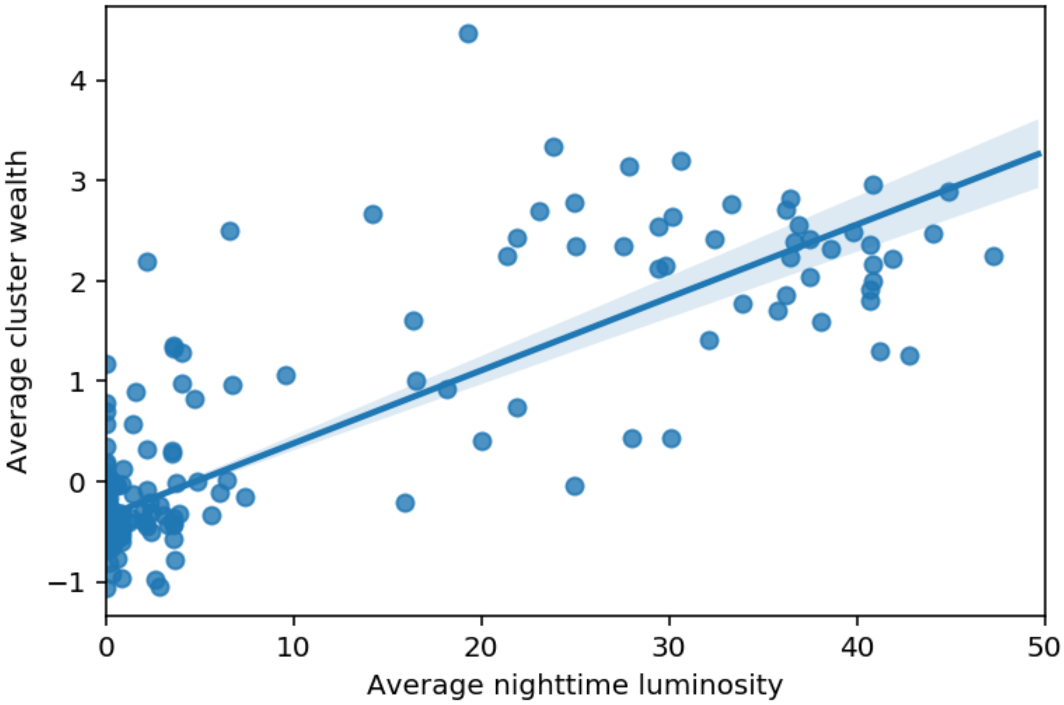


Figure 8: Regression model of cluster wealth against nighttime luminousity

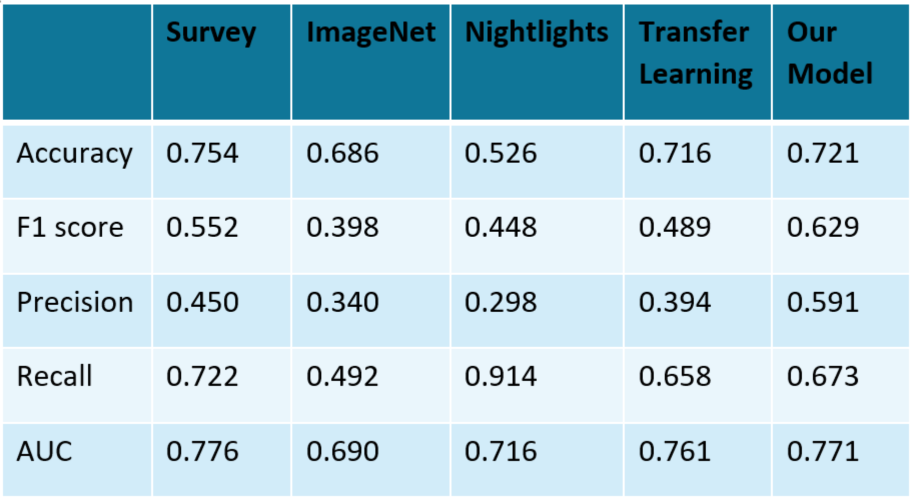
daytime imagery can predict village wealth. R-square value of the best model: 0.602 So this shows that nightlights are better indicators of poverty than daytime imagery.

### 4.4.3 Test whether features extracted by CNN can predict poverty

Now we try to predict average wealth using the features extracted from the daytime imagery using CNN. The R-sqaured value of the best model is 0.694. This is an improvement than just using basic features from daytime images. But it is still than what we achieved from the nightlights. Since nightlights data is scarce and daytime imagery is becoming increasingly available, therefore we can try to predict nightlights from the daytime images and then use that data to predict wealth.

### 4.4.4 Test whether these deep features can predict wealth

Using the predicted nightlights, we try to predict the average wealth of each cluster. R-square value =0.725 This shows that using the predicted nightlights value instead daytime features gives better estimation of poverty.



| | Survey | ImageNet | Nightlights | Transfer Learning | Our Model |
|---|---|---|---|---|---|
| Accuracy | 0.754 | 0.686 | 0.526 | 0.716 | 0.721 |
| F1 score | 0.552 | 0.398 | 0.448 | 0.489 | 0.629 |
| Precision | 0.450 | 0.340 | 0.298 | 0.394 | 0.591 |
| Recall | 0.722 | 0.492 | 0.914 | 0.658 | 0.673 |
| AUC | 0.776 | 0.690 | 0.716 | 0.761 | 0.771 |

Figure 9: Cross validation test performance for predicting aggregate level poverty measures. Survey is trained on survey data collected in the field. All other models are based on satellite imagery

# Chapter 5

# Conclusion

## 5.1 Summary

We introduce a new transfer learning approach for analyzing satellite imagery that leverages recent deep learning advances and multiple data-rich proxy tasks to learn high-level feature representations of satellite images. This knowledge is then transferred to data-poor tasks of interest in the spirit of transfer learning. We demonstrate an application of this idea in the context of poverty mapping and introduce a fully convolutional CNN model that, without explicit supervision, learns to identify complex features such as roads, urban areas, and various terrains. Using these features, we are able to approach the performance of data collected in the field for poverty estimation. Remarkably, our approach outperforms models based directly on the data-rich proxies used in our transfer learning pipeline. Our approach can easily be generalized to other remote sensing tasks and has great potential to help solve global sustainability challenges.

## 5.2 Future Work

We faced some difficulities and some of those were not solved in this work such as predicting subnational wealth for new countries, training the models on data for other countries.This task is more difficult because absolute levels of night light emissions can vary considerably across countries.One of these problems is 'overglow', or the fact that light can spill over from one cell to adjacent cells . This leads to a loss of precision in the night lights data – for example, it is difficult to detect a dark neighborhood in a city due to the spillover of light from bright neighborhoods nearby. We will try to overcome these difficulties and work to improve the accuracy of the predictions.

# References

[1] uPiaggesi, Simone, et al. ″Predicting City Poverty Using Satellite Imagery.″ Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[2] uHall, G. Brent, Neil W. Malcolm, and Joseph M. Piwowar. ″Integration of remote sensing and GIS to detect pockets of urban poverty: The case of Rosario, Argentina.″ Transactions in GIS 5.3 (2001): 235-253.

[3] uPerez, Anthony, et al. ″Poverty prediction with public landsat 7 satellite imagery and machine learning.″ arXiv preprint arXiv:1711.03654 (2017).

[4] Xie, Michael, et al. ″Transfer learning from deep features for remote sensing and poverty mapping.″ Thirtieth AAAI Conference on Artificial Intelligence. 2016.

[5] Abelson, B.; Varshney, K.; and Sun, J. 2014. Targeting direct cash transfers to the extremely poor. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 1563–1572. ACM .

[6] Chatfield, K.; Simonyan, K.; Vedaldi, A.; and Zisserman, A. 2014. Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531

[7] Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; and Darrell, T. 2013. DeCAF: A deep convolutional activation feature for generic visual recognition. CoRR abs/1310.1531.

[8] Independent Expert Advisory Group Secretariat. 2014. A world that counts: Mobilising the data revolution for sustainable development. Technical report.

[9] Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R. B.; Guadarrama, S.; and Darrell, T. 2014. Caffe: Convolutional architecture for fast feature embedding. CoRR abs/1408.5093 .

[10] Dupuis, Kate, and M. Kathleen Pichora-Fuller. ”Recognition of emotional speech for younger and older talkers: Behavioural findings from the Toronto Emotional Speech Set.” Canadian Acoustics 39.3 (2011): 182-183.

[11] Le, Q. V.; Ranzato, M.; Monga, R.; Devin, M.; Chen, K.; Corrado, G. S.; Dean, J.; and Ng, A. Y. 2012. Building high-level features using large scale unsupervised learning. In International Conference on Machine Learning .

[12] Long, J.; Shelhamer, E.; and Darrell, T. 2014. Fully convolutional networks for semantic segmentation. CoRR abs/1411.4038.

[13] Mnih, V., and Hinton, G. E. 2010. Learning to detect roads in high-resolution aerial images. In Computer Vision–ECCV 2010. Springer. 210–223.

[14] Mnih, V., and Hinton, G. E. 2012. Learning to label aerial images from noisy data. In Proceedings of the 29th International Conference on Machine Learning (ICML-12), 567– 574.

[15] Murthy, K.; Shearn, M.; Smiley, B. D.; Chau, A. H.; Levine, J.; and Robinson, D. 2014. Skysat-1: very high-resolution imagery from a small satellite. In SPIE Remote Sensing, 92411E–92411E. International Society for Optics and Photonics.

[16] Oquab, M.; Bottou, L.; Laptev, I.; and Sivic, J. 2014. Learning and transferring mid-level image representations using convolutional neural networks. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR ’14, 1717–1724. Washington, DC, USA: IEEE Computer Society .

[17] Pan, S. J., and Yang, Q. 2010. A survey on transfer learning. Knowledge and Data Engineering, IEEE Transactions on 22(10):1345–1359.

[18] Razavian, A. S.; Azizpour, H.; Sullivan, J.; and Carlsson, S. 2014. CNN features off-the-shelf: an astounding baseline for recognition. CoRR abs/1403.6382.

[19] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C.; and Fei-Fei, L. 2014. ImageNet large scale visual recognition challenge. International Journal of Computer Vision 1–42 .

[20] Varshney, K. R.; Chen, G. H.; Abelson, B.; Nowocin, K.; Sakhrani, V.; Xu, L.; and Spatocco, B. L. 2015. Targeting villages for rural development using satellite image analysis. Big Data 3(1):41–53.

[21] Wolf, R., and Platt, J. C. 1994. Postal address block location using a convolutional locator network. In Advances in Neural Information Processing Systems, 745–752. Morgan Kaufmann Publishers.

[22] World Resources Institute. 2009. Mapping a better future: How spatial analysis can benefit wetlands and reduce poverty in Uganda.

[23] Xie, M.; Jean, N.; Burke, M.; Lobell, D.; and Ermon, S. 2015. Transfer learning from deep features for remote sensing and poverty mapping. CoRR abs/1510.00098.

[24] Zhou, B.; Lapedriza, A.; Xiao, J.; Torralba, A.; and Oliva, A. 2014. Learning deep features for scene recognition using Places database. In Advances in Neural Information Processing Systems, 487–495.

[25] Beger, Andreas; Cassy L Dorff Michael D Ward (2014) Ensemble forecasting of irregular leadership change. Research Politics 1(3) (http://rap.sagepub.com/content/ 1/3/2053168014557511).

[26] Cederman, Lars-Erik; Nils B Weidmann  Nils-Christian Bormann (2015) Triangulating horizontal inequality: Toward improved conflict analysis. Journal of Peace Research 52(6): 806–821.

[27] Center for International Earth Science Information Network (CIESIN)  Centro Internacional de Agricultura Tropical (CIAT) (2005) Gridded population of the world v3 (GPWv3). Palisades, NY: CIESIN, Columbia University (http://sedac.ciesin.columbia.edu/gpw/).

[28] Chen, Xi  William D Nordhaus (2011) Using luminosity data as a proxy for economic statistics. Proceedings of the National Academy of Sciences 108(21): 8589–8594.

[29] Doll, Christopher (2008) CIESIN Thematic Guide to NightTime Light Remote Sensing and its Applications. Center for International Earth Science Information Network .

[30] Elvidge, Christopher D; Kimberley E Baugh, Eric A Kihn, Herbert W Kroehl, Ethan R Davis  Chris W Davis (1997) Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. International Journal of Remote Sensing 18(6): 1373–1379.

[31] Hegre, Håvard  Nicholas Sambanis (2006) Sensitivity analysis of empirical results on civil war onset. Journal of Conflict Resolution 50(4): 508–535.

[32] Henderson, Vernon; Adam Storeygard  David N Weil (2011) A bright idea for measuring economic growth. American Economic Review 101(3): 194–199.

[33] Hodler, Roland  Paul A Raschky (2014) Regional favoritism. Quarterly Journal of Economics 129(2): 995–1033.

[34] Jerven, Morten (2013) Poor Numbers: How We Are Misled by African Development Statistics and What To Do About It. Ithaca, NY: Cornell University Press.

[35] Kuhn, Patrick  Nils B Weidmann (2015) Unequal we fight: Between- and within-group inequality and ethnic civil war. Political Science Research and Methods 3(3): 543–568.

[36] Kyba, Christopher CM; Stefanie Garz, Helga Kuechly, Alejandro Sa nchez de Miguel, Jaime Zamorano, Ju rgen Fischer  Franz Ho lker (2014) High-resolution imagery of earth at night: New sources, opportunities and challenges. Remote Sensing 7(1): 1–23.

[37] Mellander, Charlotta; Kevin Stolarick, Zara Matheson  Jose Lobo (2013) Night-time light data: A good proxy measure for economic activity? CESIS Electronic Working Paper Series number 315. Royal Institute of Technology (https://static.sys.kth.se/itm/wp/cesis/cesiswp315.pdf).

[38] Michalopoulos, Stelios  Elias Papaioannou (2013) Pre-colonial ethnic institutions and contemporary African development. Econometrica 81(1): 113–152.

[39] Min, Brian; Kwawu Mensan Gaba, Ousmane Fall Sarr  Alsassane Agalassou (2013) Detection of rural electrification in Africa using DMSP-OLS night lights imagery. International Journal of Remote Sensing 34(22): 8118–8141.

[40] National Geophysical Data Center (2014a) DMSP-OLS nighttime lights time series, version 4 (http://ngdc.noaa. gov/eog/dmsp/-downloadV4composites.html).

[41] https://www.google.com/search?q=convolutional+picturessource