

Semi-Supervised Multitask Learning on Multispectral Satellite Images Using Wasserstein Generative Adversarial Networks (GANs) for Predicting Poverty

Anthony Perez^{*}
Stanford University
Stanford, CA - 94305
aperez8@cs.stanford.edu

Swetava Ganguli^{*}
Stanford University
Stanford, CA - 94305
swetava@cs.stanford.edu

Stefano Ermon[†]
Stanford University
Stanford, CA - 94305
ermon@cs.stanford.edu

George Azzari[‡]
Stanford University
Stanford, CA - 94305
gazzari@stanford.edu

Marshall Burke[§]
Stanford University
Stanford, CA - 94305
mburke@stanford.edu

David Lobell[¶]
Stanford University
Stanford, CA - 94305
dlobell@stanford.edu

Abstract

Obtaining reliable data describing local poverty metrics at a granularity that is informative to policy-makers requires expensive and logistically difficult surveys, particularly in the developing world. Not surprisingly, the poverty stricken regions are also the ones which have a high probability of being a war zone, have poor infrastructure and sometimes have governments that do not cooperate with internationally funded development efforts. We train a CNN on free and publicly available daytime satellite images of the African continent from Landsat 7 to build a model for predicting local economic livelihoods. Only 5% of the satellite images can be associated with labels (which are obtained from DHS Surveys) and thus a semi-supervised approach using a GAN [33], albeit with a more stable-to-train flavor of GANs called the Wasserstein GAN regularized with gradient penalty [15] is used. The method of multitask learning is employed to regularize the network and also create an end-to-end model for the prediction of multiple poverty metrics.

1. Introduction

In a recent paper [18], Convolutional Neural Networks (CNNs) were trained to extract image features from daytime satellite images with nightlight luminosity labels from all over the African continent which were then shown to

explain upto 75% of the variation in local economic livelihood indicators. This is a significant step towards using free and publicly available satellite imagery to measure the need for or the impacts of economic and social policies without requiring expensive and logistically difficult surveys. This ability would specifically benefit policy makers or philanthropic organizations to direct aid and effort in places that most need them [38], [11], [19]. If we focus on just the continent of Africa, gaps in collection of data are exacerbated by poor infrastructure, ravaging civil wars, and inadequate government resourcing. According to data from the World Bank, 39 of 59 African countries conducted fewer than two surveys during the years 2000 to 2010, from which nationally representative poverty measures could be constructed. Of these, 14 conducted no such surveys during this period [38]. Sometimes, governments do not have enough resources to conduct these surveys while others lack the motivation to highlight their lackluster performance; least of all document it [11], [34]. Clearly, closing these large data gaps for accurate measurements of the targets of the United Nations Sustainable Development Goals in every country is prohibitively expensive and institutionally difficult [19].

Recently, an increasing and consistent supply of satellite data has added accurate and sophisticated geo-spatial information into studying environmental health, economics, geopolitics, etc. [32]. The unique insight attainable from such datasets has continued to be integrated into the intelligence operations of both public (e.g. NSA, CIA) and private sector (e.g. Skybox, etc.) entities to the point that a number of modern processes would not be possible in their absence [22]. With advances in sensor technology, modern satellites are capable of capturing hyperspectral data which enables the use of satellite data in a variety of ways which were not

^{*}Equal Contributions. Enrolled in CS 231N Convolutional Neural Networks for Visual Recognition during Spring 2017 at Stanford

[†]Assistant Professor, Department of Computer Science

[‡]Research Associate, Department of Earth System Science

[§]Assistant Professor, Department of Earth System Science

[¶]Professor, Department of Earth System Science

thought to have been possible before [31]. Inspired from [18], we train CNNs to identify satellite image features that are used to predict an *asset wealth index* (AWI, as defined in [18]) that is a stable measure of a household’s long-run economic status and serves as a proxy quantitative measure for the United Nations Sustainable Development Goals in continental Africa. As part of the Sustainability and AI Lab at Stanford, we want to further this effort by taking advantage of recent progress in semi-supervised learning using Wasserstein GANs. We collected multi-spectral imagery from Landsat 7 covering the African continent ranging from 2004 to 2015 (amounting to about 22 TB). As labels, we have access to Demographic and Health Surveys (DHS) which provide the AWIs for multiple countries for the years of interest. However, the coverage of these surveys is sparse compared to the large amounts of satellite image data that we have at our disposal. Therefore, we choose to implement semi-supervised learning algorithms that can utilize large amounts of unlabelled data along with a small set of labeled data. [33] and [21] demonstrate methods of using multi-conditional likelihood objectives and shared latent variables for variational inference to boost prediction on domains where labels are scarce but data is abundant, which is the setting under which our project operates. [33] demonstrates that GANs [14] can be used to take advantage of unlabeled data and thus perform semi-supervised learning. However, GANs can be finicky and are in general difficult to train. A variant of GANs, called the Wasserstein GAN (WGAN) has been shown to be more stable to train [6]. WGANs can sometimes generate low-quality samples or fail to converge in some settings due to the use of weight clipping to enforce a Lipschitz constraint on the critic. This problem is alleviated by the use of a gradient penalty [15]. We use the WGAN-GP approach in tandem with the approach of [33] to build our semi-supervised network. In addition, [9] provides evidence that multi-task learning improves generalization of machine learning models “by using the domain information contained in the training signals of related tasks as an inductive bias”. Instead of building an individual model for each poverty metric of interest to us, we resort to multitask learning with the methods proposed in [21] by setting up each poverty metric prediction task as a classification task at the end of the discriminator network and use the available Landsat images to predict AWI across countries in Africa. This enables us to learn the correlations between these metrics as well as regularize the large capacity network. The details of this procedure are described in the sections below. The input to our model is a satellite image and the output is the predicted class in each classification task which in turn is our prediction range of a particular poverty metric. For certain metrics, policy makers seldom care about the actual value of the metric. Rather, the range is more important in policy making. For example,

“between 10 to 20 people suffer from malnutrition” is as informative as “12.83 people suffer from malnutrition”. This justifies our approach of setting up each prediction task as a classification task. The model, consisting of the generator, discriminator and each successive classification task is trained end-to-end. The prediction tasks that are of interest to us are to predict nightlight intensity, population density, distance to nearest road, land cover type, and AWI simultaneously. Of these, predicting AWI is of utmost importance. In addition to the validation accuracies obtained from the discriminator, models are also compared based on a metric measuring accuracy on predicting this score. Once the model is trained, image features are extracted from the last layer of the discriminator and a cross-validated linear model is built on top of it. A quantitative measure of the performance of our models is the Pearson Correlation Coefficient between the predicted AWI and the ground truth available from the DHS surveys. This is consistent with the approach employed in [18]. Multiple experiments, each described in subsequent sections, are performed to obtain a model that best predicts the AWI. Qualitatively, we also visualize the first layer weights from the discriminator to understand the sensitivity of the initialization schemes employed in the experiments to the prediction accuracy.

2. Satellite Images for Related Work

Use of passively collected data for predicting economic or social indicators or to study natural disasters and their local effects has been a growing trend in the research community. Sources of such data could be satellites, social media, mobile phone networks, etc. A popular recent approach uses satellite images of luminosity at night (“nightlights”) to estimate economic activity [17], [10], [23] and [29]. [25] use satellite images to predict land use while [13] propose a method for spatial and temporal analysis of life-threatening vegetation fires using satellite images that may be useful for weather departments and government agencies to create contingency plans during wildfire season. [12] and [37] use satellite images to study local economic activities and their correlation to global economic indicators like Gross Domestic Product (GDP), Ecosystem Services Product (ESP), etc. The economic and ecological costs of deforestation in Africa and South America and its effects on the local markets has been investigated from a scientific and econometrics viewpoint in [27] and [24]. As developing countries urbanize and strain the availability of resources, studying the environmental and economic effects of this rapid change is aided by temporal analysis of satellite imagery. [8] and [28] study the impact of urbanization on micro-climates of regions surrounding these cities while [35] are able to measure the land use due to urban expansion. Wall Street firms also use satellite surveillance to gather market-moving information. Private-sector compa-

nies like DigitalGlobe in Colorado and GeoEye in Virginia build and launch satellites and take pictures for US government intelligence agency clients and private-sector satellite analysis firms like Remote Sensing Metrics LLC. Finally, satellite images have also been used to track utilization of renewable energy sources by [26] and [7].

3. Description of Data and Preprocessing

As mentioned in the introduction, we collected multi-spectral imagery from Landsat 7 (Enhanced Thematic Mapper Plus (ETM+)) [2] covering the African continent ranging from 2004 to 2015 (amounting to about 22 TB). Images from this satellite have 9 spectral bands viz. (i) Blue, (ii) Green, (iii) Red, (iv) Near Infrared (NIR), (v) Shortwave Infrared (SWIR) 1, (vi) Shortwave Infrared (SWIR) 2, (vii) Thermal 1, (viii) Thermal 2 and (ix) Panchromatic. The spatial resolution of bands 1 to 6 is 30m while the resolution of bands 7 and 8 is 60m and that of the last band is 15m. All bands can collect one of two gain settings (high or low) for increased radiometric sensitivity and dynamic range, while Band 6 (Thermal) collects both high and low gain for all scenes and is thus provided as two separate bands. The approximate scene size is 170 km north-south by 183 km east-west (106 mi by 114 mi). Rather than using the Blue, Green, and Red bands as provided by Landsat 7, these visible bands are first pan-sharpened [20] [3] which results in 15m resolution versions of the color bands. Pan-sharpening has proved to be beneficial in a variety of satellite imagery tasks [20] which justifies its usage.

As labels, we have access to Demographic and Health Surveys (DHS) which provide the AWIs for multiple countries in Africa for the years of interest. However, the coverage these surveys provided is sparse compared to the large amounts of satellite image data that we have at our disposal. Figure 4 shows the locations in Africa where DHS labels are available. While sampling training data, which is a mix of labeled and unlabeled data, we experiment with two sampling strategies. One way to sample is to sample uniformly all over the African continent. Shown in figure 1 are the geographic locations of the available satellite data from Landsat that are classified as rural, semi-urban, and urban when this sampling strategy is used. The figures are slightly deceptive, as a single pixel corresponds to an area much larger than an image. The data is class balanced, with 50,000 images extracted for each class. However, this comes at a cost of overlapping in the imagery for the semi-urban and urban classes. Furthermore, very few locations that have satellite images associated with them have DHS AWI labels.

To alleviate this problem, another strategy is to restrict the distance of samples from the locations where labeled data is available. This proximity is an assurance that the labeled and unlabeled images come from nearly similar image distributions. Shown in figure 2 are the geographic locations

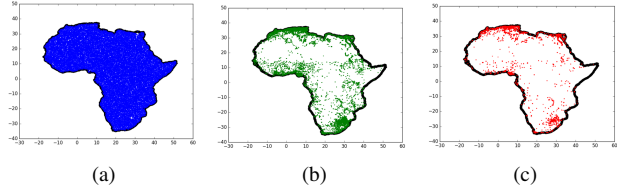


Figure 1: Geographic locations of satellite image samples which are classified as rural (class 0 in blue), semi-urban (class 1 in green), and urban (class 2 in red) when sampling is performed from the entire African continent.

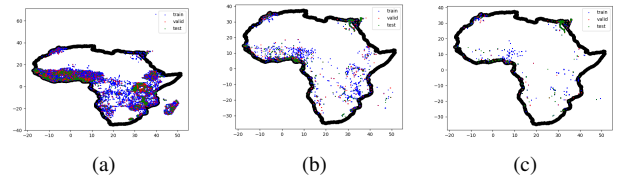


Figure 2: Geographic locations of satellite image samples which are classified as (a) rural (class 0), (b) semi-urban (class 1), and (c) urban (class 2) when the sampling is performed from regions that are near the locations where DHS survey labels are available. Within each class, the data is split as training (blue), validation (red) and test (green)

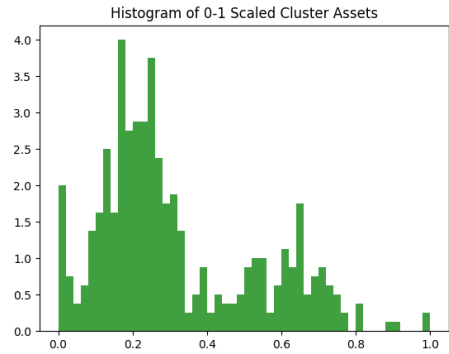


Figure 3: A histogram of scaled assets with 50 buckets demonstrating the distribution of the assets. Clearly, 50 bins is too large since many buckets are left empty. For this metric, an appropriate number of bins is 30.

of the satellite image samples when sampling is done from regions close to the locations where DHS labels are available following the second sampling strategy. Transitioning to the Africa-wide multitask setting, we have 91522 total satellite images which are bucketed into the 3 classes mentioned before i.e. rural, semi-urban and urban. As described in table 1, there are 64031 images in the rural bucket, 20050 images in the semi-urban bucket and 14639 images in the

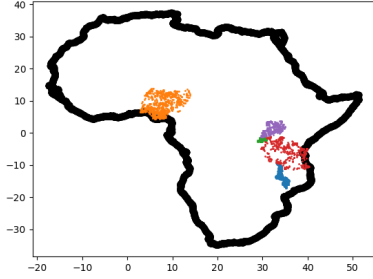


Figure 4: The locations of the most recent DHS survey in each country used in by [18]. There are 4839 locations in total.

Class	0	1	2
Train	59896	18486	13140
Validation	2805	883	788
Test	1330	681	711
Total	64031	20050	14639

Table 1: A description of the per-class distribution of the available data and a 70-20-10 train-val-test split

urban bucket. As is evident, these classes are not class-balanced. The train-val-test splits are created to ensure that they do not overlap with each other. However, there are images within the training dataset that have partial overlaps. Each task in the multitask framework is set up as a classification task. Since the final predictions for each task of interest to us corresponds to a continuous value, we bin the range of the values of each task into an appropriate number of buckets. If the number of buckets is too low, then we risk having many data points in few of the buckets which results in losing the ability to differentiate between them. If the number of buckets is too high, some buckets may be empty and have no data points representing them. As a case study, the AWI is binned into 50 buckets in figure 3. Clearly, this is the case where the number of buckets is too high and some buckets are completely empty. The appropriate number of buckets for AWI turns out to be 30.

Each modeling task performed in the multi-task training paradigm is associated with an “importance weight” that is proportional to the importance of accuracy of prediction for that task. Said differently, the weight is proportional to the cost incurred when an incorrect prediction is made for a particular learning task. For example, in our case being able to predict the nighttime intensity and AWI score correctly is more important than some of the other tasks. These importance weights are supplied from collaboration with economic and policy experts in the lab. In addition, every example in a class is weighted inversely proportional to the number of available examples in that class so that the total importance of all examples per class is the same. This

ensures that the model does not selectively do well in predicting a certain class while failing to predict another class. Thus, in effect, we calculate the product of the two aforementioned weights as a “weight-per-example-per-class-per-task”. We have 4839 locations where we have labels from the DHS survey for AWI values and 86683 locations where there are no labels for the asset values. We call this dataset the “real dataset” and distinguish the points where we have labels by calling them “labeled real data” and those without labels by calling them “unlabeled real data”. Shown in figure 5 are two images with their 9 spectral bands. The GAN

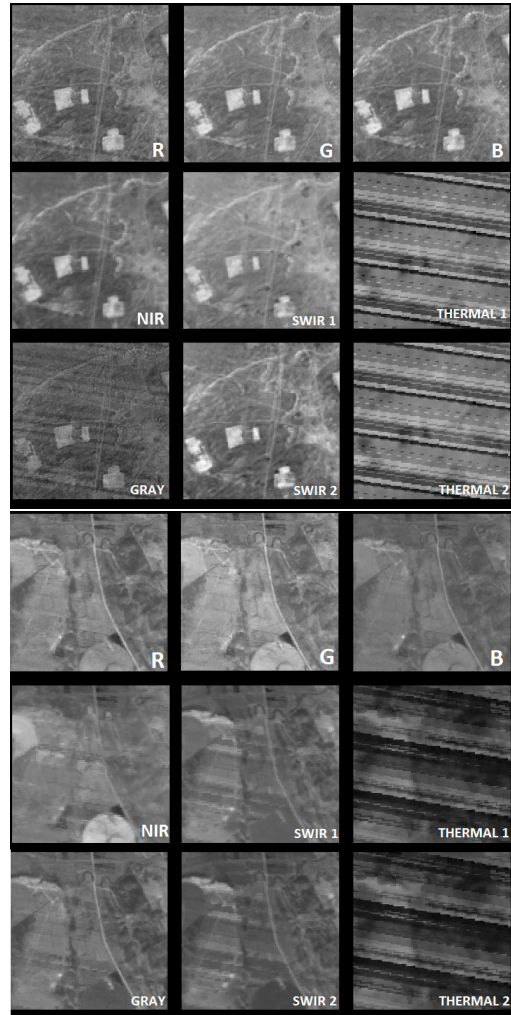


Figure 5: Two examples of actual Landsat 7 images and their 9 spectral bands.

generator supplies a fake image for every real image so that the WGAN-GP algorithm [15] can be followed. Finally, once the model is trained, the features from the last layer are extracted for the images that have ground truth labels associated with them. A linear model is built on top of these features using cross-validated ridge regression. The model

predictions are then compared to the ground truth values and a Pearson Correlation Coefficient is calculated. This is another metric by which we evaluate the performance of our models. In the work we have done so far, we have restricted our attention to the DHS survey taken in the year 2011 in Uganda. This survey contains only 400 data points corresponding to “geographic village cluster” level AWI labels.

4. Methods

As pointed out in [33], GANs [14] boost the learning power of a semi-supervised model. This is implemented by having a generator supply “fake images” and the discriminator being modified such that if the classification task is associated with K classes, the model now has to predict $(K+1)$ classes - K of them associated with real classes of interest while the $(K + 1)$ th class represents the case where the provided image is fake. We extend the loss proposed by [33] to fit into a multitask paradigm. In the multitask framework, instead of having one classification task, we have multiple tasks, each of which is predicted by the model using a fully connected layer that stems from a shared model body. Thus we append an extra class to the set of classes corresponding to each task, as shown in Figure 6. However, vanilla GANs can be unstable during training [6]. It has been shown that Wasserstein GANs [6] are a more stable variant of GANs. Thus in addition to adding an extra class to each task, an extra task is added that corresponds to the WGAN task of predicting whether the input image is real or fake. The multitask loss is the sum of each of these losses. Thus, the multitask loss function is minimized when the model (i) can predict the class association to be $(K+1)$ -th class on all tasks and predict that the image is fake in the added WGAN task, if the provided image is fake, (ii) can associate a real labeled image with the correct label for the image in each task and predict that the image is real in the added WGAN task, and (iii) provided an unlabeled real image, it predicts that the image is not in the $(K+1)$ -th class on all tasks and predicts that the image is real in the WGAN task. Here we use the word “predict” to loosely mean estimation of some property that is not directly observed, rather than its common meaning of inferring something about the future. Note that the discriminator in the WGAN task computes the Earth Mover Distance (Wasserstein-1 Distance or EMD) between the distribution of real images and the distribution of images generated by the generator and thus is not “predicting” a class in the conventional sense. Recently, [15] introduced the idea of enforcing a soft version of the 1-Lipschitz constraint required to minimize the WGAN value function constructed using the Kantorovich-Rubinstein duality. This circumvents having to use weight clipping originally proposed by [6] which was shown to be problematic [15]. For the WGAN task, we implement Algorithm 1 from [15] by computing $D(x)$ from a fully connected layer that again stems from

the shared model body.

Thus for our semi-supervised tasks, the loss function incorporates labeled and unlabeled real data and WGAN generated fake data using the Kullbeck-Liebler divergence while the WGAN unsupervised task incorporates unlabeled real data and WGAN generated fake data using the EMD. To overcome the difference in the scales of our losses, we scale the WGAN loss with the hyperparameter α so that the scales of the two contributions to the total discriminator loss are similar. Summarizing the algorithm, the loss functions for the discriminator and the generator are:

Discriminator:

\hat{x} linear $\epsilon \sim U[0, 1]$ superposition of real & fake image

$$L_{D_{\text{WGAN}}} = \mathbb{E}_{z \sim P_z} [D(G(z))] - \mathbb{E}_{x \sim P_r} [D(x)] + \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}} [(|\nabla_{\hat{x}} D(\hat{x})|_2 - 1)^2]$$

$$L_{D_t} = -\mathbb{E}_{x, y \sim P_t} [\log p_{\text{model}_t}(y|x, y < K_t + 1)] - \mathbb{E}_{x \sim P_r} [\log(1 - p_{\text{model}_t}(y = K_t + 1|x))] - \mathbb{E}_{z \sim P_z} [\log p_{\text{model}_t}(y = K_t + 1|G(z))]$$

$$L_{D_{\text{multitask}}} = \sum_{t \in \text{Tasks}} w_t L_{D_t}$$

$$L_D = \alpha L_{D_{\text{WGAN}}} + L_{D_{\text{multitask}}}$$

Generator:

$$L_{G_{\text{WGAN}}} = -\mathbb{E}_{z \sim P_z} [D(G(z))]$$

$$L_{G_{\text{multitask}}} = \sum_{t \in \text{Tasks}} w_t \mathbb{E}_{z \sim P_z} [\log p_{\text{model}_t}(y = K_t + 1|G(z))]$$

$$L_G = \alpha L_{G_{\text{WGAN}}} + L_{G_{\text{multitask}}}$$

The entire training pipeline along with which losses are used in updating and back-propagating into the model can be summarized in figure 6. The dashed lines show the loss functions that affect the model weight updates via back-propagation. The generator used is inspired from the DC-GAN deep convolutional generator [30]. The architecture we use for the generator is shown in figure 7. The output of this generator is set to be $224 \times 224 \times 9$. The discriminator is the 50-layer ResNet from [16]. More specifically, the shared model body mentioned previously is composed of all layers of the 50-layer ResNet model from [16] up to the average pooling layer. The output of the average pooling layer is fed through a fully connected layer for each task as mentioned above. Furthermore, when we refer to extracting features in the subsequent section, the features are exactly the output of the average pooling layer.

While training the discriminator, we monitor the training and validation accuracies of the semi-supervised tasks. Once the discriminator has been trained, we extract the last layer of the shared model body and use this as a feature representation of an input image. With these features, we train a linear ridge regression model on the locations where DHS survey labels for AWI are available. In this study, we

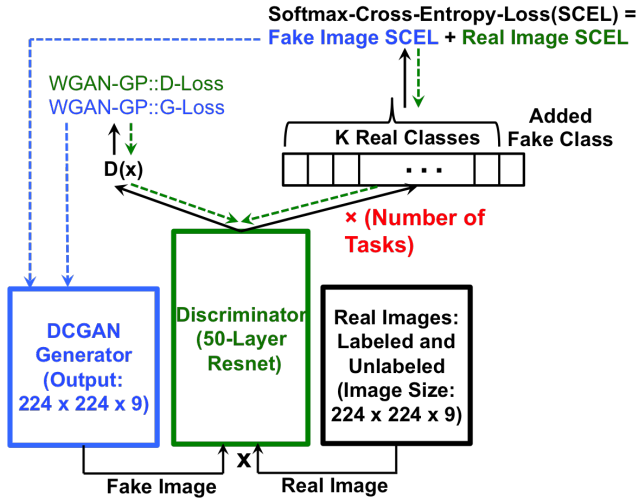


Figure 6: The WGAN Training Pipeline. The generator is a deep convolutional generator while the discriminator is a 50-layer ResNet. Solid lines show forward (inference) pass while the dashed lines show the backpropagation path.

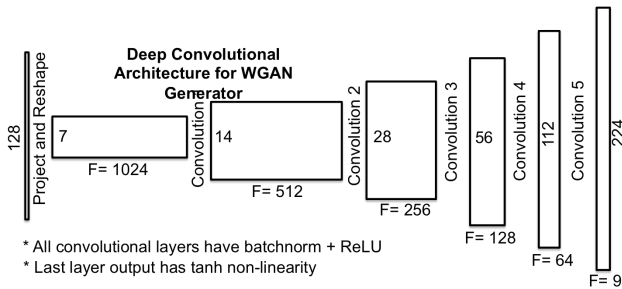


Figure 7: DCGAN [30] inspired deep convolutional generator. Each convolutional layer is followed by a batchnorm layer and a ReLU non-linearity. The last layer output has a tanh nonlinearity.

focus on the Uganda for the year 2011. The regression is performed using a doubly nested cross-validation approach. The inner loop of cross validation is used to search for the best weight on the regularization term (a hyperparameter) in ridge regression. That weight is used to make predictions for the test set in the outer loop of the doubly nested cross validation. Thus, at the end of each outer loop, a subset (specifically, the test set) of the datapoints are associated with a model prediction based on a regression over the other four folds. The predictions across all folds are shown in figure 11 for the experiments described in the next section.

5. Experiments and Results

Our efforts so far have been focused on replicating the results by [18]. To this end, we have created a pipeline that allows us to train a model based off the 50-layer ResNet

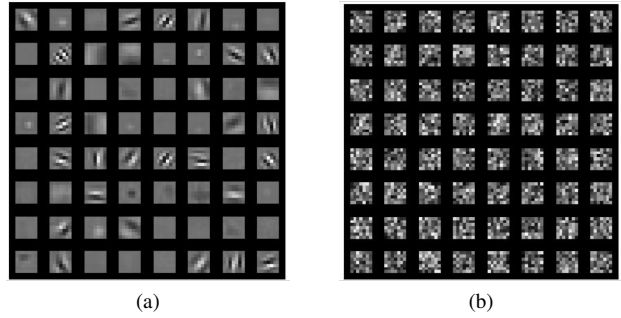


Figure 8: On the left in figure (a) are shown the 64 first layer conv filters for the NIR channel when the initialization of the weights is “same init” meaning that the weights in each channel is the mean of the RGB channels of the pre-trained ResNet weights. On the right in figure (b) are shown the same filters at convergence when the initialization of the weights is “random init” meaning that the weights are chosen from a truncated normal distribution with mean and standard deviation of the pretrained ResNet weights.

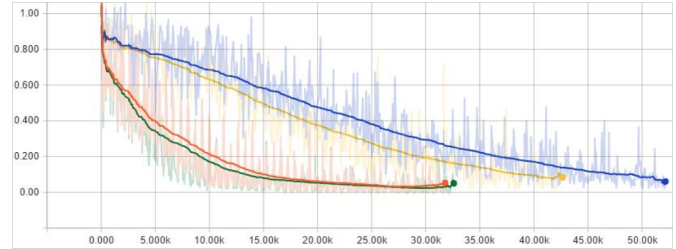


Figure 9: The plot shows the convergence behavior of the loss function of the discriminator for four experiments: (Orange) Use “Around DHS dataset” with “same init” initialization scheme; (Green) Use only the RGB channels of the “Around DHS dataset” with “same init” initialization scheme; (Blue) Use “All Africa dataset” with “random init” initialization scheme; and (Yellow) Use “All Africa dataset” with “same init” initialization scheme.

described in [16] to predict binned night light categories. This is followed by a process that extracts features from the ResNet model for each location corresponding to a cluster in the Uganda DHS survey. Finally we train a simple linear regression model from those features to predict the AWI score. In order to replicate the transfer learning approach used by [18], we have divided night lights into the same bins. The primary differences between our work and the work in [18] are the locations of the training data imagery, the model used, and the number of wavelengths of light captured in the imagery.

We have conducted several experiments regarding the high level hyperparameters of our model which include the

initialization of our model, the dataset our model is trained with, and the number of Landsat 7 channels supplied to the model. Note that because our imagery is multispectral (and thus has 9 channels), the size of the filters in the first layer of our ResNet model have changed relative to their Imagenet counterparts. As such, initialization directly from Imagenet parameters is not possible for the network. We define two initialization schemes, both of which initialize all weights in the second layer through the last layer to the corresponding Imagenet weights. The first initialization scheme, which we denote “random init”, initializes the RGB components of the filters in the first layer to their Imagenet values, and then initializes the other channel’s components randomly using a truncated normal distribution with the mean and standard deviation of the weights in the first layer of the ResNet model pretrained on Imagenet . The second initialization scheme, which we denote “same init”, again initializes the RGB components of the filters in the first layer to their Imagenet values, but now initializes the other channel’s components with the mean of the pretrained Imagenet weights taken across the RGB axis. In other words, for every position in each first layer filter, its six non-RGB channel weights are initialized as the mean of the three RGB channel weights available from the pretrained Imagenet model. Additionally, we define two datasets over which we conduct experiments. These datasets are described in more detail in Section 3. The first dataset is denoted the “All Africa dataset” and was sampled uniformly from an evenly spaced grid over the African continent. This dataset is visualized in Figure 1. The second dataset is denoted the “Around DHS dataset” and was sampled such that its location distribution is similar to the location distribution from which the DHS surveys are sampled. This dataset is visualized in Figure 2. Finally we experiment with whether including only the Blue, Green, and Red channels rather than all nine Landsat 7 channels degrades or improves the performance of the model.

A manual hyperparameter search was done over the learning rate, batch size, and learning rate decay. Figure 9 shows the training loss of each of our experiments using their final hyperparameter settings. All models use a batch size of 115, which is the maximum our hardware would allow. Additionally, all models feature l_2 weight decay on all layers with a regularization strength of 0.00004. Our best model uses a learning rate of 0.01 and a learning rate decay of 0.98 per epoch. We additionally lower the learning rate by a factor of 5 after 25 epochs of training as this is where the model plateaus. The best performing model is initialized using the “same init” initialization scheme, uses only the RGB channels from the Landsat 7 imagery, and is trained on the “Around DHS dataset”.

We evaluate our methods on the AWI available in DHS survey sampled in Uganda in the year 2011. First, predic-

Name	DHS Dataset	Same Init	RGB Only	Assets r^2	Training Accuracy	Validation Accuracy
All Africa Random Init	X	X	X	0.51	96%	66%
All Africa Same Init	X	✓	X	0.52	98%	70%
DHS Same Init	✓	✓	X	0.53	98%	65%
DHS Same Init RGB Only	✓	✓	✓	0.57	97%	68%
Nightlights	N/A	N/A	N/A	0.484	100%	100%
Jean et al.	N/A	N/A	N/A	0.66	N/A	N/A

Figure 10: A table summarizing the results of our experiments. Names in the first four rows refer to experiment qualities. Nightlights refers to using the raw nightlight values. Jean et al. refers to [18]. The first two columns highlight the initialization scheme and the dataset used. The third column denotes if all the 9 channels were used or only the first 3 channels were used. The fourth column denotes the r^2 obtained in the linear model when the features extracted by the discriminator are used. The last two columns are the training accuracies and validation accuracies for the discriminator models.

tions are made for all locations with data available using the ridge regression model trained as described in Section 4. We then evaluate the success of our model by computing the Pearson correlation coefficient between the predicted AWI values and the actual AWI values. Reporting the Pearson correlation coefficient allows for a fair comparison between this work and the work in [18]. As a tool to understand our methods, the night light training and validation prediction accuracies are reported as well.

Figure 10 reports the results of the experiments conducted while Figure 11 compares each model’s predicted AWI to the ground truth data in the 2011 DHS survey in Uganda. By comparing each model’s night light prediction accuracy on the training and validation sets we can see signs of overfitting on the training data, as the training accuracy is much higher than the validation accuracy. While regularization would be beneficial, incorporating additional tasks may also be sufficient to reduce overfitting and would likely result in improved generalization of the features extracted by the model [9]. Sadly we have not surpassed the results of our “oracle” [18], who achieved a correlation coefficient of approximately 0.66 (we have achieved 0.57). However, even with overfitting we do surpass the baseline correlation coefficient computed between night light values and the AWI, which is 0.484.

As one may expect, using the “Around DHS dataset” and the “same init” initialization resulted in increase model performance. Using the “Around DHS dataset” shifts the

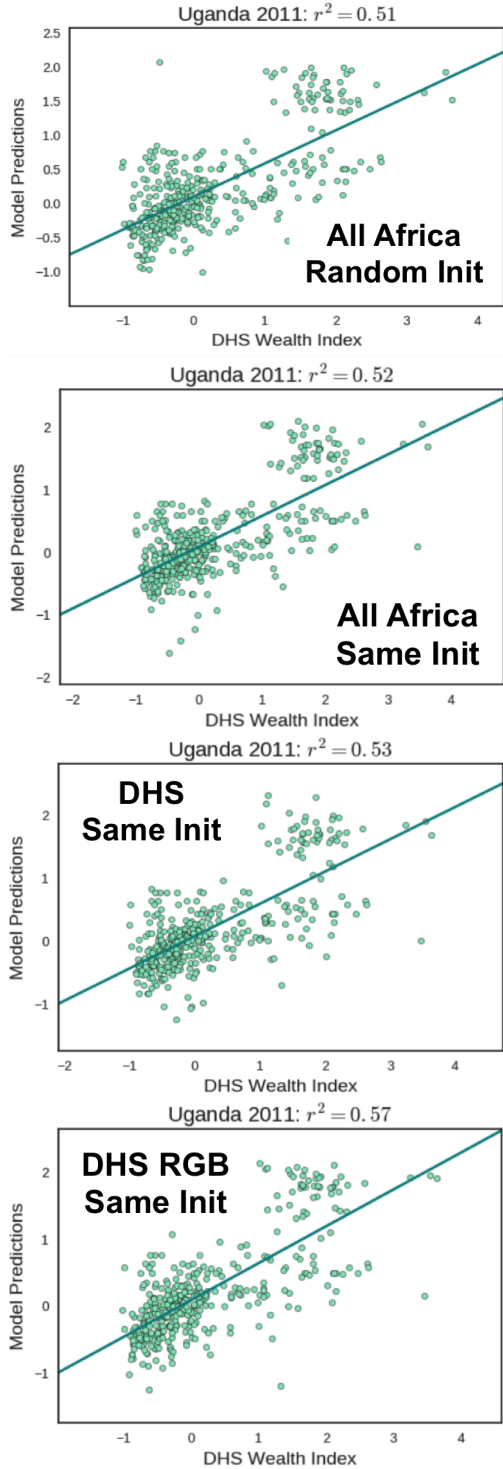


Figure 11: Figures show the relation between the model predictions and the ground truth values of AWI in Uganda for the year 2011. The line demonstrates the linear ridge regression model that is built on top of the predictions. Asset values are predicted using the doubly nested cross validation strategy described in Section 4.

model’s training distribution closer to the distribution of images in which DHS AWI assets are available which reduces the problem of covariate shift [36]. Likewise, using the “same init” allows the model to retain the information present in the pretrained weights and may better pair the non-RGB filter weights with the pretrained RGB weights being used in both initialization schemes. In order to better understand the effects of the different initialization schemes, the weights of the first layer filters corresponding to the Near Infrared (NIR) channel are visualized in Figure 8. It seems that when the “same init” is used, the model retains filters that look similar to those of the pretrained ImageNet model. Unexpectedly, removing the non-visible channels from the input imagery results in a better performing model. This may be explained by the model’s propensity to overfit, as the model may be fitting to noise in the additional image channels. We expect that when overfitting is resolved, either by adding additional tasks or incorporating stronger regularization, that using all Landsat 7 image bands would result in a more powerful model.

6. Conclusions and Future Work

We have proposed a technique to extend the semi-supervised learning method proposed in [33] to a multitask setting. This is done by adding an extra class corresponding to fake images to each semi-supervised task. We also incorporate the WGAN loss proposed in [15] in order to achieve greater stability in training our generator. This involves adding an additional task whose loss is described in Algorithm 1 in [15]. We operate in a domain where only 5% of the data is labeled while 95% of the data is unlabeled. In domains where labels are scarce, the multitask WGAN model proposed can be used for simultaneously predicting the output of several semi-supervised tasks. As of now, the discriminator is capable of over-fitting to the data which demonstrates that the model has sufficient capacity. As more tasks are added, they should have a regularizing effect on the model and improve the generalization of our model’s features which is one of our next steps. Our model’s predictions comfortably achieve a higher correlation coefficient with ground truth AWI than raw night lights do. We consider outperforming night lights to be a reasonable “baseline” for the model’s success. We are still a small distance away from overcoming the 0.66 correlation coefficient achieved by [18] which is our “oracle”. The WGAN generator has proven very hard to train. Training the generator to convergence and getting fake images that mimic actual satellite images is our next goal.

7. Acknowledgements

We would like to thank Stefano Ermon, Volodymyr Kuleshov¹, Marshall Burke, and David Lobell for providing algorithmic and domain insights as well as constructive criticism. We would like to thank George Azzari and Matthew Davis² for helping us gather and understand the satellite and economic data. We would like to thank Neal Jean³ for both fruitful discussion as well as providing the code used to train the ridge regression model from features extracted by our Resnet-based model to predict poverty. This assisted in allowing us to compare our results accurately. We would like to thank [4] for providing starting code for our Resnet architecture as well as [15] for providing the code from which our DCGAN generator was derived. Likewise, we would like to thank [33] for making public the code on which our implementation of their semi-supervised loss is based. Our work is implemented using Tensorflow [5] and our data was gathered using Google Earth Engine [1].

References

- [1] Google earth engine. <https://earthengine.google.com/>. [Online; accessed 17-October-2016].
- [2] Landsat 7: Description of Spectral Bands. <https://landsat.usgs.gov>. [Online; accessed 8-June-2017].
- [3] Panchromatic Image Sharpening of Landsat 7 ETM+. <https://landsat.usgs.gov/panchromatic-image-sharpening-landsat-7-etm>. [Online; accessed 9-June-2017].
- [4] Resnet in tensorflow. <https://github.com/ry/tensorflow-resnet/blob/master/resnet.py>. [Online; accessed 17-October-2016].
- [5] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [6] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. *arxiv:1701.07875v2 [stat.ML]*, 2017.
- [7] A. W. Azhari, K. Sopian, A. Zaharim, and M. Al Ghoul. A new approach for predicting solar radiation in tropical environment using satellite images-case study of malaysia. *WSEAS Transactions on Environment and Development*, 4(4):373–378, 2008.
- [8] T. N. Carlson and S. T. Arthur. The impact of land use/land cover changes due to urbanization on surface microclimate and hydrology: a satellite perspective. *Global and planetary change*, 25(1):49–65, 2000.
- [9] R. Caruana. Multitask learning. *Machine Learning*, 1997.
- [10] Chen. *Proc. Natl. Acad. Sci. U.S.A.* 108, 85898594, 2011.
- [11] Devarajan. *Rev. Income Wealth* 59, S9S15, 2013.
- [12] C. N. Doll, J.-P. Muller, and J. G. Morley. Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, 57(1):75–92, 2006.
- [13] E. Dwyer, J.-M. Grégoire, and J.-P. Malingreau. A global analysis of vegetation fires using satellite images: Spatial and temporal dynamics. *Ambio (Sweden)*, 1998.
- [14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *arXiv:1406.2661v1 [stat.ML]*, 2014.
- [15] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of wasserstein gans. *arXiv:1704.00028 [cs.LG]*, 2017.
- [16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
- [17] Henderson. *Am. Econ. Rev.* 102, 9941028, 2012.
- [18] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 2016.
- [19] Jerven. Poor numbers: How we are misled by african development statistics and what to do about it. *Cornell Univ. Press*, 2013.
- [20] K. Kpalma, M. Chikr El-Mezouar, and N. Taleb. Recent Trends in Satellite Image Pan-sharpening techniques. In *1st International Conference on Electrical, Electronic and Computing Engineering*, Vrnjacka Banja, Serbia, June 2014.
- [21] V. Kuleshov and S. Ermon. Hybrid deep discriminative/generative models for semi-supervised learning. *Submitted to UAI*, 2017.
- [22] Meyer. Silicon Valley’s new spy satellites. *The Atlantic*, 2014.
- [23] Michalopoulos. *Q. J. Econ.* 129, 151213, 2013.
- [24] C. Muñoz-Piña, A. Guevara, J. M. Torres, and J. Braña. Paying for the hydrological services of mexico’s forests: Analysis, negotiations and results. *Ecological economics*, 65(4):725–736, 2008.
- [25] G. C. Nelson and D. Hellerstein. Do roads cause deforestation? using satellite images in econometric analysis of land use. *American Journal of Agricultural Economics*, 79(1):80–88, 1997.
- [26] I. Nygaard, K. Rasmussen, J. Badger, T. T. Nielsen, L. B. Hansen, S. Stisen, S. Larsen, A. Mariko, and I. Togola. Using modeling, satellite images and existing global datasets for rapid preliminary assessments of renewable energy resources: The case of mali. *Renewable and Sustainable Energy Reviews*, 14(8):2359–2371, 2010.
- [27] O. O. Omo-Irabor, S. B. Olobaniyi, J. Akunna, V. Venus, J. M. Maina, and C. Paradzayi. Mangrove vulnerability modelling in parts of western niger delta, nigeria using satellite images, gis techniques and spatial multi-criteria analysis (smca). *Environmental monitoring and assessment*, 178(1-4):39–51, 2011.

¹Ph.D. Candidate, Department of Compute Science

²Center on Food Security and the Environment, Stanford University, Stanford, CA, USA.

³Ph.D. Candidate, Department of Electrical Engineering

- [28] T. Owen, T. Carlson, and R. Gillies. An assessment of satellite remotely-sensed land cover parameters in quantitatively describing the climatic effect of urbanization. *International journal of remote sensing*, 19(9):1663–1681, 1998.
- [29] Pinkovskiy. *Q. J. Econ.* 131, 2016.
- [30] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, abs/1511.06434, 2015.
- [31] A. Romero, C. Gatta, and G. Camps-Valls. Unsupervised deep feature extraction of hyperspectral images. In *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing (WHISPERS)*. IEEE, volume 6, 2014.
- [32] Safyan. Overview of the planet labs constellation of earth imaging satellites. *ITU Symposium/Workshop, Prague*, 2015.
- [33] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. Improved techniques for training gans. *arXiv:1606.03498 [cs.LG]*, 2016.
- [34] Sandefur. Benefits and costs of the data for development targets for the post-2015 development agenda. *J. Dev. Stud.* 51, 116132, 2015.
- [35] K. C. Seto, M. Fragkias, B. Güneralp, and M. K. Reilly. A meta-analysis of global urban land expansion. *PloS one*, 6(8):e23777, 2011.
- [36] H. Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227 – 244, 2000.
- [37] P. C. Sutton and R. Costanza. Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecological Economics*, 41(3):509–527, 2002.
- [38] World Bank. Povcalnet online poverty analysis tool, <http://iresearch.worldbank.org/povcalnet/>. 2015.