# **Predicting asset ownership in Africa using satellite imagery**

Nicolas Suarez<sup>1</sup> Othman Bensouda<sup>2</sup> Edoardo Yin<sup>3</sup> 1 nsuarez@stanford.edu

2 othmanb@stanford.edu



## **Introduction**

- We predicted **asset ownership in Africa** making use of Landsat-8 satellite imagery and ground truth data from the Demographic and Health Surveys (DHS).
- We trained a **ResNet-18 model** (He et al., [2016\)](#page-0-0), a sub-class of a **Convolutional Neural Network (CNN)**, and used it to predict asset ownership levels over 6.72 km by 6.72 km patches of land in Africa. **Summary**
- We used **transfer learning** to train our model to perform a **regression task**, initializing our model with weights originally used to classify images. We predicte as the control of Landsate making use of Landsatellite in Africa making use of Landsatellite in an<br>Satellite image of Landsatellite in and ground truth data satellite in and ground truth data satellite in and
- The performance of our best model is much better than the performance of our baseline models, but our model might be overfitting, and the predictions it generates are far off from our ground truth data. We obtained lower  $R^2$  validation set values than Yeh et al. [\(2020\)](#page-0-1), a paper that attempts our same task.

## **Ground truth: Demographic and Health Surveys (DHS) surveys**

- We combined all the household DHS surveys for African countries that have been published since 2014 that have a matching file with the geocoordinates of each cluster.
- Our variable of interest is the score in the wealth index obtained by each household. This score is produced using Principal Component Analysis over survey questions, including but not limited to access to drinkable water, sewerage, electricity, and ownership of farming and non-farming assets.
- We geocoded 12,511 locations across 24 countries, combining data from 26 DHS surveys between 2014 and 2021. access to drink able water, sewerage, sewerage, electricity, and our farming and non-farming and non-farming a
- In Panel (a) of Figure [1](#page-0-2) we show the spatial distribution of our ground truth data.  $f$  Figure 1,500 chaw the spatial distribution of our ground truth data  $2010 - 1022$

- We retrieved imagery using the Google Earth Engine API. For each DHS cluster with a geocoded location, we defined a patch of  $6.72 \text{ km} \times 6.72 \text{ km}$  centered in our location, and we retrieved an image for the patch using the Landsat 8 Surface Reflectance Tier 1 Collection. **Satellite images: Landsat-8 Surface Reflectance collection**
- We used 3 bands from this collection: Red, Green and Blue surface reflectance. We preprocessed each of our images by adding a cloud mask per pixel and then computing the per pixel and band mean composite of all the available images for the year when the DHS cluster was surveyed. patch using the Landsat o Suriace Reflectance Tier I Collection.<br>The center in our location of the center and we retrieve and we retrieve and we retrieve and we retrieve and in nds from this collection: Red, Green and Blue surface reflectance. We preprocesse
- We retrieved images for [1](#page-0-2)2,426 locations across all Africa. In Figure Panels  $(\mathrm{b})$  and  $(\mathrm{c})$  of Figure 1 we show some examples of the cloud masked imagery we produced.

Table 1. Training, validation and test set sizes

## **Satellite images: Landsat-8 Surface Reflectance collection** In Panel (a) of Figure **??** we show the spatial distribution of our ground truth data.







<span id="page-0-2"></span>

**Figure 1.** Ground truth spatial distribution and examples of images

**(a)** Location of DHS clusters in Africa

## **Production asset of a set of a friend satellite in Africa using satellite in Africa using satellite in a set of a Models**

**(c)** Image from Angola

- esidual Network, a particular kind of a Convolutional Neural Network (CNN). to initialize our model, and we adapt its architecture so we can use it for a We trained a Residual Network, a particular kind of a Convolutional Neural Network (CNN). We use transfer learning to initialize our model, and we adapt its architecture so we can use it for a regression task.
- IC DY SLACKING **RESIGUAL DIOCKS**, DIOCKS OF CONVOL stacked convolutional layers at the end (see example in Figure [2\)](#page-0-3). Residual blocks should help a deep CNN to avoid the performance problems associated with very deep networks, so if some of the final layers are not helping the model performance, their weights will be set to 0 and the block will become an identity mapping. Nullam eu nulla eleifendum nec eu lorem. Vivamus felis velit, volution en elementity mapping. **Residual Network (ResNet)**: Type of CNN with a special architecture defined by He et al. [\(2016\)](#page-0-0). These kind of models are build by stacking **Residual Blocks**, blocks of convolutional layers connected with activation functions, where the input of the block is then added to the output of the

 $\Gamma$ isis actor in metus.  $\Gamma$ **Figure 2.** Example of a residual block from He et al.(2016)



<span id="page-0-3"></span>

 $F_{\text{reinad}}$  <u>DecNet 18 model</u> a Besidual Network. Diutional layers grouped in o residual blocks connected by ReLU activation functions. The last of the original model is a linear layer connected to a softmax layer that classifies images into 1,000 classes, so we modify the last linear layer so now it outputs only one number that will represent owners layer of the original model is a linear layer connected to a softmax layer that classifies images into our predicted asset ownership, and we measure our loss as our root mean squared error (RMSE). We initialize our model using the original ResNet-18 pretrained weights. In Figure [3](#page-0-4) we show our model's<br>———————————————————— **Our implementation**: We trained a ResNet-18 model, a Residual Network that contains 18 convolutional layers grouped in 8 residual blocks connected by ReLU activation functions. The last architecture:

Figure 3. Residual Network architecture

 $\epsilon$ dictum or consected the statis seminary separation, and the statistic consection interests to are the performance of our moder against them. **dels:** We trained linear regression, Lasso regression and Ridge regres **Benchmark models:** We trained linear regression, Lasso regression and Ridge regression models to compare the performance of our model against them.

- and **benchmark models**.
- validation and test sets.
- LASSO regression performs equally bad on the training, validation and test sets. Our best model is a **ResNet-18 model** trained with:
- 200 epochs
- nullam non est elit su su su cursus est elit . Maecenas portions est elit . Maecenas portugues augustas august<br>Elit de la curmontation transformations annied to images • Training batch size of 500 images
- 4 data augmentation transformations applied to images
- Adam optimizer **Learning rate**  $= 0.001$
- 

s<br>e last **Country average asset ownership for pixels in our test set**: Our results are off from the ground nu chaen averages are negative, but we predict on ave truth. Most of the ground truth averages are negative, but we predict on average positive values of asset ownership.



the diamed a need to the arists with Editate 6 satemet magery to preated accele currenting in<br>Africa We modified the original ResNet-18 architecture so we could train it for a regression ta and we used **transfer learning** to take advantage of the pretrained ResNet-18 weights. We trained a **ResNet-18** model with Landsat-8 satellite imagery to predict **asset ownership in Africa**. We modified the original ResNet-18 architecture so we could train it for a **regression** task,

Und we ased cransier rearning to take advantage of the pretrained resiver to weights. ques and<br>• The performance of our best model is much better than the performance of our baseline models, but • • The performance of our baseline models, but the model might be overfitting, and the predictions it generates are off from our ground truth data.

time. Then with the thight morne expanding can cample one by morning more years ment one increasing the regularization in our model, adding Future work on this might include expanding our sample size by including more years with DHS

ren, C., Ferez, A., Driscon, A., Azzan, G., Tang, Z., Loben, D., Lrmon, S., & Durke, M. (2020). Osing publiciy avand<br>satellite imagery and deep learning to understand economic well-being in africa. *Nature communications*, posed of a shortcut branch and a shortcut branch and a few deeper branches. The shortcut branches. The shortcu<br>In the shortcut branches. The shortcut branches are shortcut branches. The shortcut branches. The shortcut bra Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available



$$
\left.\begin{array}{c}\n\mathbf{x} \\
\text{identity}\n\end{array}\right.
$$





<span id="page-0-4"></span>

#### **Sed consequat** id ante vel efficitur. Praesent congue massa sed est scelerisque, elementum mollis perform *identity* mapping, and their outputs are added to the outputs of the stacked layers (Fig. 2). In the stacked lay **Experiments**

In sed est finibus, vulputate nunc gravida, pulvinar lorem. In maximus nunc dolor, sed auctor eros porttitor quis. hyperparameters of our models, we experimented using data aug modifying several components of our model. cut connectare of our models we experimented using data augmentation To tune the hyperparameters of our models, we experimented using data augmentation techniques and

#### **Sed luctus, elit sit amet** dictum maximus, diam dolor faucibus purus, sed lobortis justo erat id augme<br>. end-to-end-**Data augmentation techniques**:

- acion cechinques.<br>Jont vertical and berizental flips to our images with certain probabilities Applying independent vertical and horizontal flips to our images with certain probabilities.
- reprising independent vertical and nonzontal mps to our midges with certain probabilities.<br>Rotating our images in degrees multiples of 90 (0, 90, 180 and 270) with certain probabilities. ure<br>ing different degre
- Applying different degrees of Gaussian Blurring to our images.

#### $\frac{1}{3}$  to tuneu.<br>Show the degradation problem and evaluate our probl **Hyperparameters tuned**: **Number of epochs for training.**

- **Mini-batch training size.**
- method. We show that in the show that is not contained to the state of the stat Number of frozen convolutional layers at the bottom of the model (last layers of the model).
- stic Gradient Descent or Adam). Exhibit higher training exhibit higher training exhibit  $\mathcal{L}$ Optimizer (Stochastic Gradient Descent or Adam).
- **Learning rate.**
- For Stochastic Gradient Descent, we tuned the momentum parameter (momentum makes our gradient a moving average of our previous gradients).

Similar phenomena are also shown on the CIFAR-10 set



<sup>3</sup>edoyin@stanford.edu





#### **References** codels to a set of the conditioning conditions of the representation of  $\epsilon$ sent and sent and sent residual vectors between two scales. **Concluding remarks and future work**

- $\frac{1}{2}$   $\frac{1}{2}$  trained a **PosNot 18** model with Landsat 8
- ne perform  $S<sub>1</sub>$  and the moderningite be overnicing, and the prediction We obtained lower  $R^2$  validation set values than Yeh et al. [\(2020\)](#page-0-1).
- $\overline{a}$  Euture work on this might include expanding our dropout probabilities to our nodes and tuning that hyperparameter.

<span id="page-0-1"></span>Concurrent with our work, "highway networks" [42, 43]

<span id="page-0-0"></span>He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference* ving<br>
on computer vision and pattern recognition, 770–778.

### diate layers are directly connected to auxiliary classifiers for addressing vanishing vanishing vanishing gradients. The papers of the papers of the papers of the papers o<br>The papers of the papers o **References**

## **Results**

## We computed the  $\mathbf{RMSE}$  and  $\mathbf{R^2}$  coefficient in our  $\mathbf{training}$ ,  $\mathbf{validation}$  and  $\mathbf{test}$  sets for our  $\mathbf{CNN}$

Linear regression and Ridge regression have good performance on the training set but a very poor performance on the

**A block contained with:**  $\mathbf{A} \mathbf{B}$ 

## The best model has an excellent performance on the training set, and a good performance on the validation and test sets, suggesting **overfitting** despite using **regularization**.

 $T<sub>ch</sub>l<sub>2</sub>, D<sub>em</sub>formence$ Performance metrics for baseline models

Figure 4. Average asset ownership per country for test set observations