Predicting asset ownership in Africa using satellite imagery

Nicolas Suarez [1] Othman Bensouda [2] Edoardo Yin [3]

[1]nsuarez@stanford.edu [2]othmanb@stanford.edu [3]edoyin@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), 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.
- We used **transfer learning** to train our model to perform a **regression task**, initializing our model with weights originally used to classify images.
- 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), 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.
- In Panel (a) of Figure 1 we show the spatial distribution of our ground truth data.

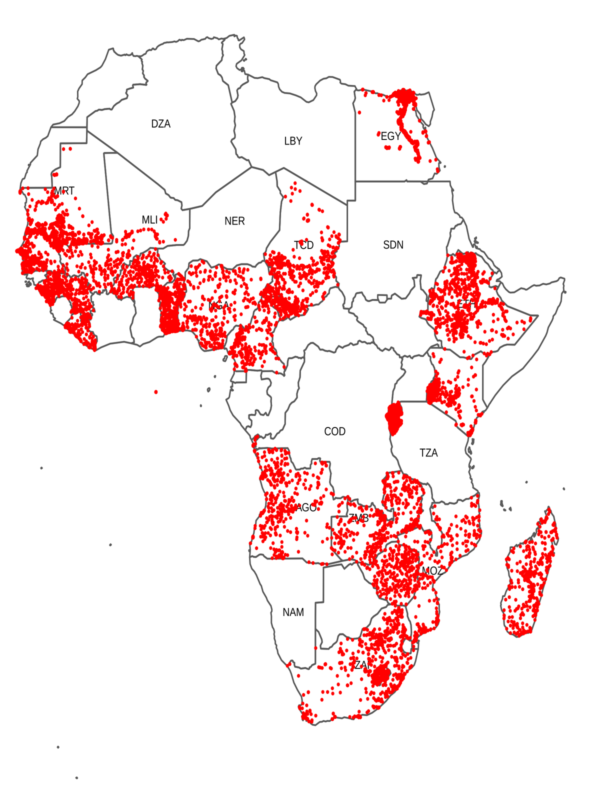
Satellite images: Landsat-8 Surface Reflectance collection

- We retrieved imagery using the Google Earth Engine API. For each DHS cluster with a geocoded location, we defined a patch of 6.72 km × 6.72 km centered in our location, and we retrieved an image for the patch using the Landsat 8 Surface Reflectance Tier 1 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.
- We retrieved images for 12,426 locations across all Africa. In Figure Panels (b) and (c) of Figure 1 we show some examples of the cloud masked imagery we produced.

Table 1. Training, validation and test set sizes

Set	Training	Validation	Test
Observations	9,941	1,243	1,242

Figure 1. Ground truth spatial distribution and examples of images



(a) Location of DHS clusters in Africa

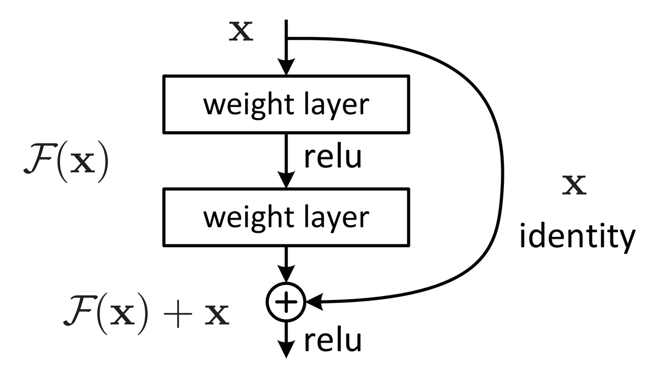
(b) Image from Kenya

(c) Image from Angola

Models

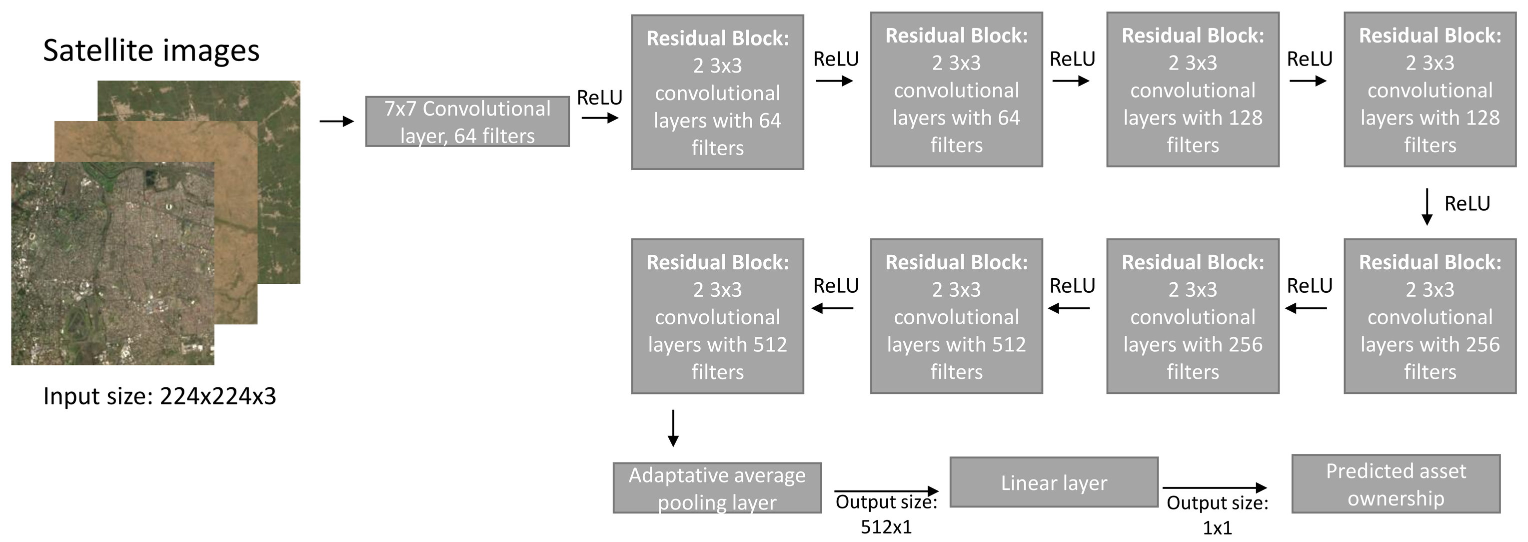
- We trained a Residual Network, a particular kind of a Convolutional Neural Network (CNN). We use transfer learning to initialize our model, and we adapt its architecture so we can use it for a regression task.
- **Residual Network (ResNet)**: Type of CNN with a special architecture defined by He et al. (2016). 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 stacked convolutional layers at the end (see example in Figure 2). 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.

Figure 2. Example of a residual block from He et al.(2016)



- **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 layer 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 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 we show our model's architecture:

Figure 3. Residual Network architecture



- **Benchmark models:** We trained linear regression, Lasso regression and Ridge regression models to compare the performance of our model against them.

Experiments

To tune the hyperparameters of our models, we experimented using data augmentation techniques and modifying several components of our model.

- **Data augmentation techniques**:
  - Applying independent vertical and horizontal flips to our images with certain probabilities.
  - Rotating our images in degrees multiples of 90 (0, 90, 180 and 270) with certain probabilities.
  - Applying different degrees of Gaussian Blurring to our images.
- **Hyperparameters tuned**:
  - Number of epochs for training.
  - Mini-batch training size.
  - Number of frozen convolutional layers at the bottom of the model (last layers of the model).
  - 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).

Results

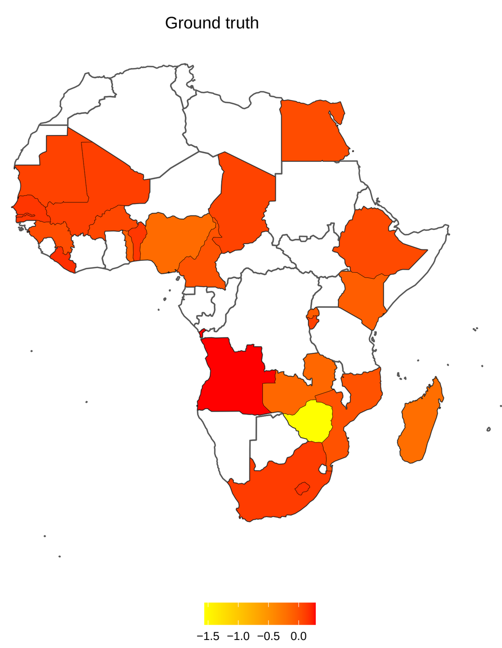
- We computed the **RMSE** and $\mathbf{R^2}$ coefficient in our **training**, **validation** and **test** sets for our **CNN** and **benchmark models**.
  - Linear regression and Ridge regression have good performance on the training set but a very poor performance on the validation and test sets.
  - LASSO regression performs equally bad on the training, validation and test sets.
- Our best model is a **ResNet-18 model** trained with:
  - 200 epochs
  - Training batch size of 500 images
  - 4 data augmentation transformations applied to images
  - Adam optimizer
  - Learning rate = 0.001
- 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**.

Table 2. Performance metrics for baseline models

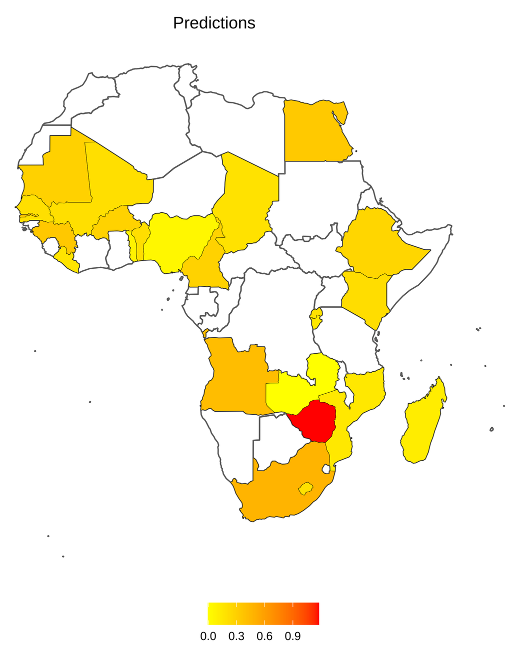
Set	Training		Validation		Test	
Metric	**RMSE**	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$
Panel A. Benchmark models						
Linear regression	0.008	1.000	2.352	-0.734	2.403	-0.683
Lasso regression	1.733	0.000	1.786	0.000	1.853	-0.001
Ridge regression	0.071	0.998	2.310	-0.673	2.359	-0.622
Panel B. Best experiment						
Residual network	0.478	0.948	0.976	0.591	1.309	0.501

Country average asset ownership for pixels in our test set: Our results are off from the ground truth. Most of the ground truth averages are negative, but we predict on average positive values of asset ownership.

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



(a) Ground truth country averages



(b) Predicted country averages

Concluding remarks and future work

- 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, and we used **transfer learning** to take advantage of the pretrained ResNet-18 weights.
- The performance of our best model is much better than the performance of our baseline models, but the model might be overfitting, and the predictions it generates are off from our ground truth data. We obtained lower R^2 validation set values than Yeh et al. (2020).
- Future work on this might include expanding our sample size by including more years with DHS surveys, using more channels of our satellite images, increasing the regularization in our model, adding dropout probabilities to our nodes and tuning that hyperparameter.

References

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature communications*, *11*(1), 1–11.