Optimal Transport For Domain Adaptation on African Satellite Imagery

By Luke Lloyd, Priyanka Nagasamudra *UC San Diego*

15 March 2023

ABSTRACT

Unsupervised domain adaptation is an important problem in machine learning. Machine learning models trained on a certain dataset often are tested on datasets that are drawn from a different distribution. The aim of our work is to find methods to provide the same performance of the machine learning model on a test distribution as the train distribution. Optimal Transport is the new We propose using optimal transport to achieve this and demonstrate results on a model used to asses poverty in Africa.

1. Introduction

Many machine learning models do not perform as well when they are deployed for real world use as they do in training. This disparity affects proper evaluation of the model and often leads models thought to perform well and deployed for use in actuality performs quite poorly. Many approaches have been used to solve this issue including data augmentation, techniques on preventing overfitting, *e.t.c.*. These methods work only under the assumption that the training data for the model is representative of the data the model would be deployed on, *i.e.* drawn from the same distribution. However in many areas, machine learning models are tested on data that is a different distribution of data than the one trained on. This issue comes under domain adaptation. We specifically look at the case of unsupervised domain adaptation where the labels of the deployed distribution is unknown.

1.1. Unsupervised Domain Adaptation

Given a train set $X_s = \{x_i^s\}_{i=1}^{N_s}$ where $x_i^s \in \mathbf{R}^d$ and a set of labels $Y_s = \{y_i\}_{i=1}^{N_s}$ where $y_i \in C$, C is the set of all labels, unsupervised domain adaption infers the labels of the test set $Y_t = \{y_i\}_{i=1}^{N_t}$ from $X_t = \{x_i^t\}_{i=1}^{N_t}$. Unsupervised domain adaption also supposes that the probability distributions $\mathbf{P_s}(x^s, y)$ and $\mathbf{P}(x^t, y)$ are not equal and are drawn from a source domain $\Omega_s \in \mathbf{R}^d$ and target domain $\Omega_t \in \mathbf{R}^d$ respectively, $\Omega_s \neq \Omega_t$.

1.2. Optimal Transport for Unsupervised Domain Adaptation

To use optimal transport for unsupervised domain adaptation, we suppose there exists some non-linear transformation

$$T : \Omega_s \to \Omega_t$$

We let $\mathbf{P}(y|x^s) = \mathbf{P}(y|T(x^t))$ where T is the most efficient transportation with respect to a given cost. We will concretize this using optimal transport.

Let r, c be probability vectors in the simplex $\Sigma_{N_s \times N_t} = \{x \in \mathbf{R}_+^{N_s \times N_t}\}$ such that $r = \sum_{i=1}^{N_s} p_i^s \delta_{x_i^s}$ and $c = \sum_{i=1}^{N_t} p_i^t \delta_{x_i^t}$ where δ_{x_i} is the Dirac Delta function at the sample x_i and p_i is the associated probability mass such that $\sum_{i=1}^{N} p_i = 1$. Define $U(r,c)$ as the transport polytope between r and c where

$$U(r,c) = \{T \in \mathbf{R}_+^{N_s \times N_t} | T\mathbf{1}_{N_s} = r, T^\top \mathbf{1}_{N_t} = c\}$$

Then, using Sinkhorn Transport algorithm as described in [1], we can find T by

$$T = \underset{T \in U(r,c)}{\arg\min} \langle M, T \rangle - \frac{1}{\lambda} h(T) \tag{1}$$

Where $M(i,j) = ||x_i^s - x_j^t||_2^2$ such that M is a cost matrix on squared Euclidean distance, $h(T) = -\sum_{i=1}^{N_s} \sum_{j=1}^{N_t} p_{ij} \log(p_{ij})$ and $p_{ij} = T(i,j)$ and $\langle \cdot, \cdot \rangle$ is the Forbineus dot product. The regularization term is added to reduce the sparsity of the transport plan T allowing for a smoother transport overall.

2. Experiment

To test optimal transport in domain adaptation, we implement Sinkhorn Transport for an unsupervised domain adaptation problem. We want to test the following pipeline:

Train:

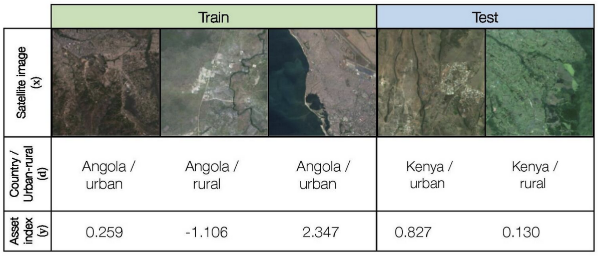
$$(X_s, Y_s) \to \text{ML Model}$$

Test:

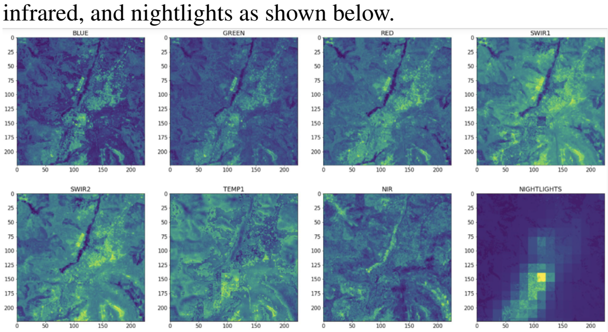
$$(X_t) \to T(X_t) \to \text{ML Model} \to \text{Prediction} Y_t$$

2.1. Dataset

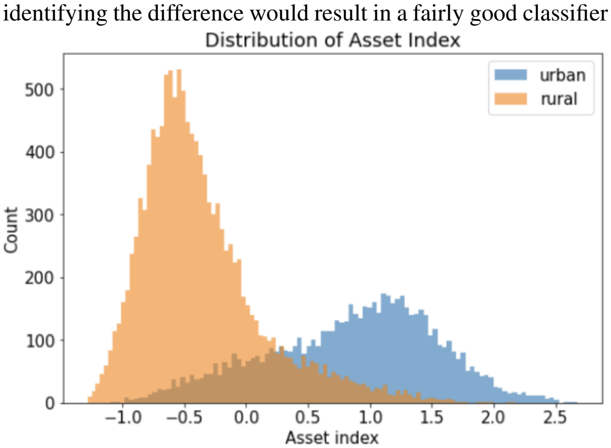
We use the poverty map dataset from WILDS [2]. This data provides satellite imagery of Africa with corresponding wealth indices used to predict the wealth of the corresponding satellite image. The images are divided by country, and by whether the image is of a urban or rural area of the county.



Each satellite image has 8 channels which are blue, green, red, short wave infrared 1, short wave infrared 2, temperature, near infrared, and nightlights as shown below.

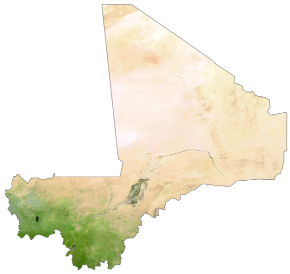


The distribution of wealth in the urban and rural areas are very different per country. which made it important to train models separately on the two distributions as otherwise simply identifying the difference would result in a fairly good classifier

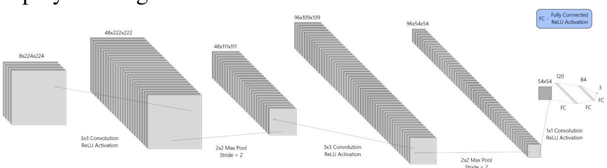


2.2. Problem Formulation

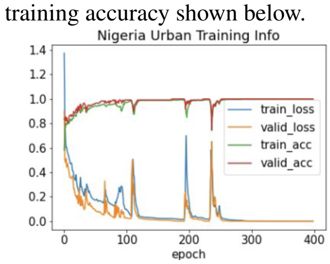
We let the source distribution be the satellite images from the country Nigeria and the target distribution to be images of Mali. We chose these two countries for the color shift between them. Mali is far more desert and therefore more brown with satellite imagery while Nigeria has far more grassland and jungle resulting in more greener satellite imagery. Mali is shown above and Nigeria is shown below.

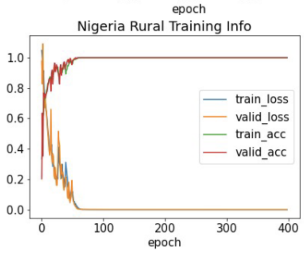


We use convolutional neural network (CNN) as our machine learning model that we train on satellite images of Nigeria and deploy on images of Mali. The architecture is shown below.



Due to the change in distribution for urban and rural wealth, we train two models – one on urban satellite images of Nigeria and the other on rural. The model is trained for 400 epochs with training accuracy shown below.





In this case, let X_s be the set of satellite images of urban or rural Nigeria, Y_s be the wealth labels where $y_i^s \in C = \{-1, 0, 1\}$. The labels are classified by the lower third of the wealth, middle third of wealth and upper third of wealth corresponding to -1, 0 and 1 respectively.

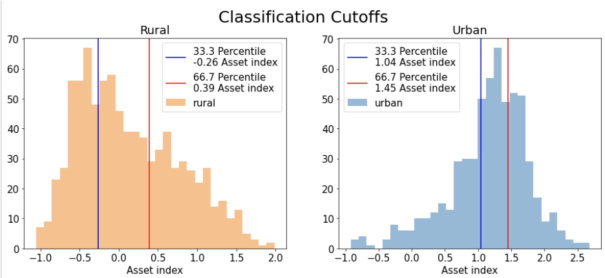
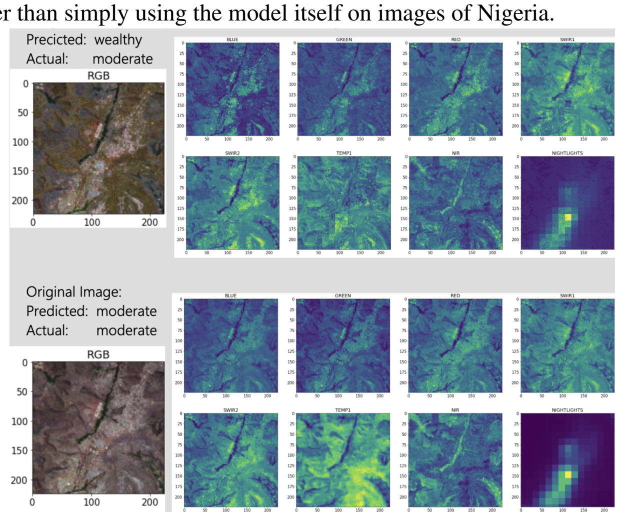


Figure 1: Shows urban and rural classification cutoffs when using Nigeria as target domain.

We then implement optimal transport by using satellite images of Mali as X_t. As Sinkhorn Transport is still quite slow for training on large amounts of data, we train only on 500 points chosen from urban (or rural) Nigeria and urban (or rural) Mali so that $N_s = N_t = 500$. We then find the transport matrix T as defined above and then transport each pixel in X_t by finding its nearest neighbor in squared Euclidean distance and using the corresponding transport plan of that pixel.

2.3. Results

We found in this case that Sinkhorn Transport did not work better than simply using the model itself on images of Nigeria.



3. Conclusion

While optimal transport did not work for unsupervised domain adaptation in this case, we believe this is due to a lack of data in our optimal transport leading to artifacts.

References

[1] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transportation distances," 2013. DOI: `10.48550/ARXIV.1306.0895`. [Online]. Available: `https://arxiv.org/abs/1306.0895`.

[2] C. Yeh, A. Perez, A. Driscoll, *et al.*, "Using publicly available satellite imagery and deep learning to understand economic well-being in africa," *Nature Communications*, 2020.