Bridges To Prosperity: Geospatial Deep Learning for Remote Site Identification

Linh Tran[1*], Mohsin Nadaf[2], Rachel Bushinsky[2], Phaneendra Sai Alluru[2], Deborshi Goswami[2], Abbie Noriega[3], James Shirley[3, †]

[1]Autodesk AI Lab [2] Autodesk [3] Bridges To Prosperity

Abstract

Infrastructure is the backbone of any country – it is essential to create employment opportunities, improve the quality of life for the poor and boost economic growth. However, nearly a billion rural people worldwide lack basic accessible transport. Building Bridges Non-Profit (BBNP) is a non-profit organization that creates rural access by working with national and district-level governments to build trailbridges that can be used by pedestrians, motorcyclists, and livestock. While effective and arguably necessary for building trailbridges at a larger scale, using a field-based method to identify where trailbridges are needed is too time- and resource-intensive to be completed at scale. This work presents a deep learning-based approach to remotely identifying river crossings that need a trailbridge using multi-modal geospatial data. First, we describe how to extract each geospatial dataset, rasterize, tile, and augment it. Then, we formulate the remote site identification problem as a binary classification problem and show how to apply supervised and semi-supervised deep learning. Our evaluation shows a performance of over 85% when testing it within a country and over 77% accuracy when transferring the model knowledge to a different country, resulting in a 61% and 34% increase in performance compared to using satellite imagery.

Introduction

Nearly a billion rural people throughout the world lack basic transportation access to essential destinations such as markets, farms, schools, financial institutions, and hospitals (Roberts, KC, and Rastogi 2006). Pedestrian ways, motorcycle tracks, and roads that are not regularly accessible present major obstacles to fundamental livelihood opportunities for rural households. These households make up 79 percent of the world's poor, with a poverty rate three times higher than in urban areas (Nations 2019b). Recently, the United Nations noted that networked infrastructure supports 72 percent of the 169 Sustainable Development Goals, highlighting the critical role that transportation infrastructure plays in ending poverty (Nations 2019a).

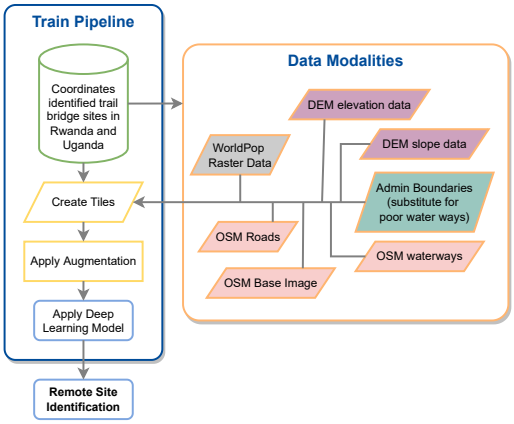


Figure 1: **Remote site identification.** Overview of deep learning based approach. Based on the data provided by Building Bridges Non-Profit (BBNP)[1], coordinates of trailbridge sites are used to extract different data modalities to create tiles. These tiles are being used for learning deep learning models to identify potential trailbridge sites.

Building Bridges Non-Profit (BBNP)[1] is a non-profit organization that creates rural access by working with national and district-level governments to build trailbridges that can be used by pedestrians, motorcyclists, and livestock. Building Bridges Non-Profit "trailbridges" are pedestrian bridges designed to be durable, cost-effective, and can be constructed in remote locations with little to no heavy machinery; they are located primarily in rural areas along "trails" or walking paths. To date, BBNP has created new safe access for an estimated 1.5 million people in 20 nations by constructing more than 400 trailbridges. A randomized controlled trial conducted on BBNP projects in Nicaragua demonstrated that communities with new trailbridges saw a 75 percent increase in farm profits, a 36 percent increase in labor market income, and a 60 percent increase in women entering the labor market (Brooks and Donovan 2021). A matched cohort study in Rwanda, which served as a pilot for a larger randomized controlled trial currently underway, built on that work and found a 25 percent increase in labor

*Work was done while at Autodesk AI Lab.

†Correspondence to kyleshirley@bridgestoprosperity.org

Keywords: rural access, trailbridge, trail bridge rural infrastructure, needs assessment, pedestrian, bridge, GIS, geospatial deep learning

Figure 2: One of the over thousands trailbridges built by BBNP in Rwanda. This bridge connects Minazi and Mataba Sectors, spans 52 meters, and serves an estimated 2,860 individuals.

market income following trailbridge construction (Thomas et al. 2020).

BBNP conducts needs assessments to determine the scope and distribution of trailbridge need. The field-based approach to needs assessment relies on relationships with government officials and community focus groups, who provide locations where bridges are needed, in addition to context and details regarding the challenges faced by rural communities in the target geography. While effective and arguably necessary for a large-scale trailbridge program, the field-based method is time- and resource-intensive and not practical for early-stage planning and development. Therefore, BBNP has been researching ways to apply machine learning to identify river crossings that require a trailbridge. As a baseline study, Building Bridges Non-Profit ran a machine learning analysis using "off-the-shelf" satellite imagery at various zoom levels to identify potential trailbridge sites. The results were underwhelming, with prominent striping in "yes" and "no" reliable predictions. A majority of research works has focused on applying deep learning to satellite imagery. However, the variance in data collection and the image quality hamper the applicability and ultimately reduce performance. In this work, we deviate from using satellite imagery and present a deep learning-based framework that leverages multiple geospatial data modalities to estimate potential trailbridge sites. In particular, our contributions are:

1. We investigate different types of geospatial data (OpenStreetMap base map and layers, population, administrative bounds and digital surface models data) and describe how to extract and rasterize them to be an appropriate input for deep learning models.
2. We propose and evaluate random tile sampling and data augmentation techniques for different geospatial data modalities.
3. We present a large-scale evaluation of our supervised and semi-supervised deep learning approaches with data for Rwanda and Uganda, and analyze the efficacy of using non-satellite-imagery geospatial data for remote site identification.

A visualization of our approach is in Figure 1.

Related Work

Geospatial data describe objects, events, or other features with a location on or near the earth's surface. Geospatial data are critical for several applications, among others, precision agriculture (Praveen and Sharma 2020; Nash, Korduan, and Bill 2009), urban planning (Sun and Du 2017; Long and Shen 2015), disaster monitoring and response (Manfré et al. 2012; Kawasaki, Berman, and Guan 2013), and climate change assessment (Konisky, Hughes, and Kaylor 2016; Hassani, Huang, and Silva 2019). Recent works have focused on applying deep learning to satellite imagery, given the success of deep neural networks in similar computer vision tasks and the sheer volume of remotely sensed imagery available. We refer to (Kiwelekar et al. 2020) for an in-depth review of deep learning for geospatial data. However, the variance in data collection and the image quality due to seasonality, clouds, haze, and sun angle differences hamper the applicability and ultimately reduce performance.

Most research approaches are limited to satellite imagery, i.e., RGB or multi-spectral images. A few works have only investigated the usage of heterogeneous data modalities. It was explored through end-to-end deep networks in (Audebert, Saux, and Lefèvre 2016; Ojogbane et al. 2021; Brown et al. 2018; Audebert, Le Saux, and Lefèvre 2018; Hu et al. 2017) for LiDAR, SAR and RGB data. While all these works investigated data fusion of various sensors, they did not study the inclusion of highly processed, semantically richer data. Only (Audebert, Le Saux, and Lefèvre 2017) used OpenStreetMap in combination with satellite imagery for semantic maps. To the best of our knowledge, the heterogenous data fusion of multiple highly-processed geospatial data modalities without satellite imagery for deep learning has not been investigated so far.

Methodology

We investigated seven datasets to identify river crossings: **administrative boundaries** provided by GADM[1] (for Rwanda, Uganda), the National Institute of Statistics Rwanda (for Rwanda) and public ArcGIS data[2] (for Uganda), **terrain elevation and slope** from the NASADEM digital elevation model (Crippen et al. 2016) hosted by the Google Earth Engine (Gorelick et al. 2017), **OpenStreetMap (OSM)** base map images, roads, and waterways (Bennett 2010), and finally, the Worldpop **population distribution** data[3] (Tatem 2017). Some datasets' relation to river crossings is more evident than others, and some datasets were included based on the assumption that they would not be informative. Most of the geospatial data modalities are provided in geospatial vector data format. We converted the vector data to a rasterized format for deep learning models to use the data for the whole countries of Rwanda and Uganda. We refer to the Appendix for a detailed description of each data modality's data extraction and preprocessing as well as a justification about the choice of data modalities.

[1] https://gadm.org/

[2] https://www.arcgis.com/home/user.html?user=BarbaraSj

[3] https://www.worldpop.org

Data Tiling

We experimented with tile sizes of 300, 600, and 1200 meters for the data tiles. With a margin of 25 meters, river crossings can lie anywhere in the tile and do not have to align with the tile center point. The margin was introduced to avoid river crossings too close to the border of the tile. We visualized the data tiles with their margin and the area in which bridge sites can lie in Figure 3a.

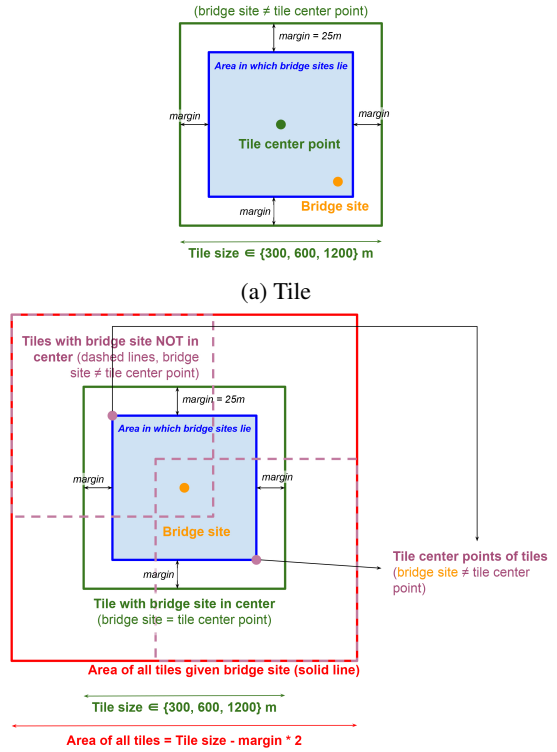
For the training and evaluation, we sample random data tiles with a potential trailbridge, i.e., need for a river crossing ("positive" tile) or without a crossing ("negative tile"). We define the squared data tiles with their center point and tile size (the square length). Given the location of a trailbridge site, we can create indefinitely many tiles that include the trailbridge site. Starting from a data tile where the tile center point and river crossing are in the same location, we can create a valid tile with the bridge site by shifting the center point. The only constraint with the shift is that both sides of the river need to reside within the valid area. The resulting area in which there are only "positive" tiles, i.e., tiles with at least one bridge site, is visualized in Figure 3b. In this area, there cannot be any center points of "negative" tiles, i.e., tiles with no bridge sites. We can construct a "negative" tile by sampling a point outside the "positive" tile area, as long as the "negative" tile does not overlap with any "positive" tile.

Data Augmentation

In the case of this project, the amount of data is proportionally small. However, the kind of data used has high complexity, i.e., the data cannot be easily learned by a linear predictor. For better generalization, we augment the training data so that our model can learn from additional synthetically modified data and is robust against different conditions, e.g., missing data, different orientations, and locations. Data augmentations for natural images have been widely researched. However, these data augmentations cannot all be used for all data modalities in this work.

The data augmentation consists of two steps. First, we augment all data modalities together. We refer to the concatenation of all data modalities as "tile image". We shift the tile center point by adding a small noise to the coordinates. With a probability of $p = 0.5$, the tile image is flipped horizontally. The image is rotated by a uniformly sampled angle in the range of $[-30, 30]$. Then, for each data modality, we developed a different type of augmentation:

- **Augmenting binary data (roads, waterways, admin bounds)**: Binary tiles are tiles with only two unique pixel values, 0 and 1. The pixel value 1 usually denotes the presence of a road or waterway or admin bound. With a probability of $p = 0.25$, a pixel with a value 1 will be set to 0. The pixel probabilities are independent of each other.
- **Augmenting terrain data (slope, elevation)**: We augment each unique terrain value by a rounded, randomly Gaussian distributed value $\mathcal{N}(0, 2)$. The same terrain value gets the same random number. We sort the unique terrain values, and we sort the random numbers. Thus, smaller terrain values get a smaller change than larger terrain values. This way, we ensure that the relative order of terrain values will not be changed through augmentation.
- **Augmenting population data**: For each pixel representing a population count, we augment this count by sampling from a Gaussian distribution and rounding this random number. Formally, given a pixel value p we augment it by $p = p + \text{round}(z), z \sim \mathcal{N}(0, 3)$.
- **Augmenting maps (OSM image)**: We augment the pixel values by sampling from a Gaussian distribution $\mathcal{N}(0, 3)$ and round this random number. In order to ensure the pixel values are between 0 and 255, we clip any values outside the range to 0 or 255.



(a) Tile

(b) Area of all possible tiles given a bridge site

Figure 3: **Visualization of data tile with a bridge site.** We assume that data tiles can contain a bridge site within a certain margin. However, the bridge site does not have to coincide with the tile center point. Given a bridge site, we can create indefinitely many tiles within the area of all tiles (outer square with red solid line).

Remote Bridge Site Identification

Notation. Assume we have a dataset $\mathcal{D} = \{\boldsymbol{X}, \boldsymbol{Y}\} \cup \{\boldsymbol{X}_{\mathcal{U}}\}$. $\boldsymbol{X}$ denotes the input, i.e., N tiles of M geospatial data modalities as input $\boldsymbol{X} \in \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N\}$, $\boldsymbol{X}_i \in \mathbb{R}^{k \times k \times M}$, where k is the tile size and M is the number of data modalities. $\boldsymbol{Y}$ represents the corresponding binary labels $\boldsymbol{Y} \in \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\} \in [0, 1]$. In addition to the la-

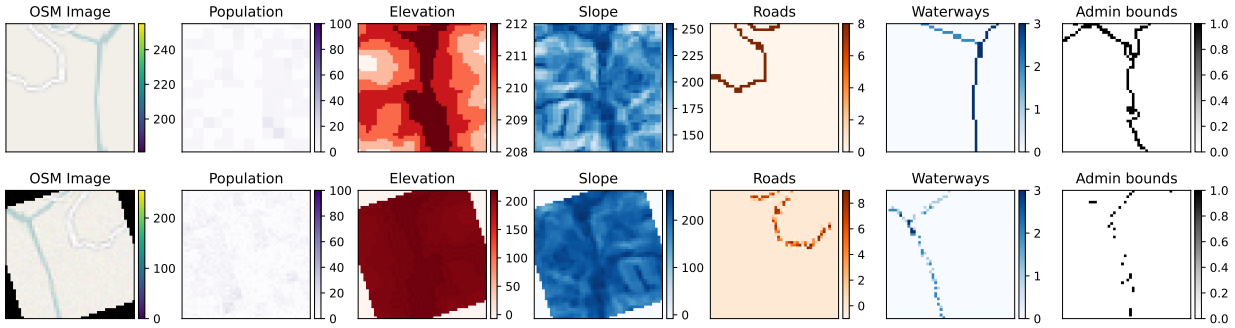


Figure 4: **Exemplary data augmentation of geospatial data modalities.** The first row show all geospatial data modalities of one location. The second row shows the tile after applying a random sequence of augmentation.

belled dataset, we have an unlabelled dataset $\{\boldsymbol{X}^{\mathcal{U}}\} = \{\boldsymbol{X}_1^{\mathcal{U}}, \ldots, \boldsymbol{X}_K^{\mathcal{U}}\}, K \gg N$.

Remote site identification as binary classification. We formalize the problem of identifying a potential remote bridge site as a binary classification problem. Given an input $\boldsymbol{X}_i \in \boldsymbol{X}$ and label $\boldsymbol{y}_i$, we first augment the input $\hat{\boldsymbol{X}}_i = \text{Augment}(\boldsymbol{X}_i)$ and use a neural network to output logits, the unnormalized final scores of your model. We calculate p_i, the normalizated prediction output, by applying the Softmax function to the logits. In our evaluation, the model is trained by minimizing a binary cross-entropy loss

$$\mathcal{L}(\hat{\boldsymbol{X}}_i, \boldsymbol{y}_i) = -\boldsymbol{y}_i \log p_i + (1 - \boldsymbol{y}_i) \log(1 - \log p_i). \quad (1)$$

Semi-supervised learning. Even though BBNP has been extensively collecting data for the remote site assessment, the data size remains rather small scaled, i.e., several thousand data points. In case of an over-parameterized neural network, optimization with small datasets can result in overfitting and thus, poor generalization. Semi-supervised learning has proven to be an effective technique for leveraging unlabelled data. We apply MixMatch (Berthelot et al. 2019), a state-of-the-art semi-supervised approach, to learning and classifying remote sites.

In order for the paper to be self-contained, we briefly describe the MixMatch (Berthelot et al. 2019) approach and its loss objective. Given an unlabelled sample $\boldsymbol{X}_i^{\mathcal{U}}, i = 1, \ldots, K$, we create two augmentations $\hat{\boldsymbol{X}}_{i,1}^{\mathcal{U}} = \text{Augment}(\boldsymbol{X}_i^{\mathcal{U}})$ and $\hat{\boldsymbol{X}}_{i,2}^{\mathcal{U}} = \text{Augment}(\boldsymbol{X}_i^{\mathcal{U}})$. We use the neural network f to "guess" the class distributions for the augmented sample $\hat{p}_{i,1}^{\mathcal{U}} = f(\hat{\boldsymbol{X}}_{i,1}^{\mathcal{U}})$ and $\hat{p}_{i,2}^{\mathcal{U}} = f(\hat{\boldsymbol{X}}_{i,2}^{\mathcal{U}})$. We average the predictions over both two augmentations and sharpen it with a temperature T and denote the output as $\hat{p}_i^{\mathcal{U}} = \text{Sharpen}(\text{Average}(\hat{p}_{i,1}^{\mathcal{U}}, \hat{p}_{i,2}^{\mathcal{U}}), T)$. We concatenate both labelled and unlabelled sets $\mathcal{W} = \{\hat{\boldsymbol{X}}_i, \boldsymbol{y}_i\}_{i=1,\ldots,N} \cup \{\hat{\boldsymbol{X}}_{i,j}^{\mathcal{U}}, \hat{p}_i^{\mathcal{U}}\}_{i=1,\ldots,K;j=1,2}$ and shuffle it. We use $\mathcal{W}$ to augment $\boldsymbol{X}^{\mathcal{U}}$ and $\hat{\boldsymbol{X}}$ with MixMatch (Zhang et al. 2018), i.e., literally mixing up the input and its corresponding label. The resulting outputs are $\{\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}}\}$ and $\{\tilde{\boldsymbol{X}}^{\mathcal{U}}, \boldsymbol{P}^{\mathcal{U}}\}$. The loss consists of two parts

$$\mathcal{L}_{\text{MixMatch}} = \underbrace{\mathcal{L}(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}})}_{\text{labelled loss}} + \underbrace{\mathcal{L}(\tilde{\boldsymbol{X}}^{\mathcal{U}}, \boldsymbol{P}^{\mathcal{U}})}_{\text{unlabelled loss}}, \quad (2)$$

where $\mathcal{L}$ is the binary cross-entropy loss. Further MixMatch uses a mean teacher(Tarvainen and Valpola 2017), i.e., predicting with exponential moving average of the model. We refer to the original paper of MixMatch (Berthelot et al. 2019) for a deep-dive into the method.

Inference with averaging over several tiles. Instead of using just one tile to predict whether a location is a remote site, we used several randomly sampled tiles to average the prediction to obtain a more robust prediction. Given a location with latitude and longitude. x, y, we can sample H data tiles $\boldsymbol{X}_i, \ldots, \boldsymbol{X}_H$. The resulting prediction is calculated by averaging over the normalized predictions $p_i, \ldots, p_H$

$$y = \arg\max\left(\frac{1}{H}\sum_i^H f(\boldsymbol{X}_i)\right) = \arg\max\left(\frac{1}{H}\sum_i^H p_i\right). \quad (3)$$

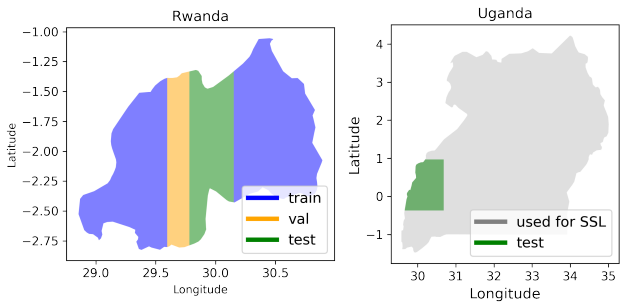
Evaluation

We evaluate the performance of using geospatial data to predict remote sites that require a trailbridge. In what follows, we describe experimental settings, our main results and ablation studies to understand the approach and data modalities.

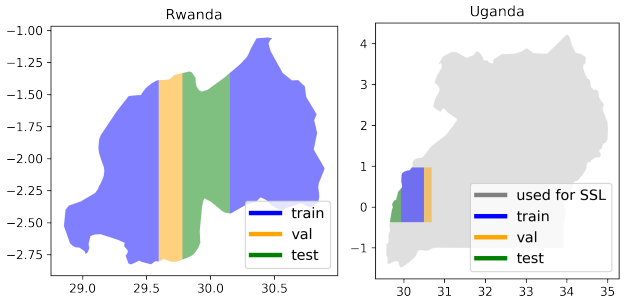
Experimental Settings

Data Collection Over four years, Building Bridges Non-Profit identified 1,321 sites in Rwanda and 247 sites in the three districts in Uganda which were used for this project. All 1,568 sites were visited in person to validate the site and record GPS coordinates.

Data Split for Train, Validation and Testing We split the data into three sets for training and evaluating the models: train, validation, and test. We used the train set for optimizing the model parameters, the validation set for model selection ("choose the model with the best validation accuracy"),



(a) Data version 1 (*in- and out-domain test*)



(b) Data version 2 (*in-domain test*)

Figure 5: **Geographical split of Rwanda and Uganda data into train, validation and test.** We split the labelled data according to its geographic location into different sets (train, validation and test). The unlabelled Uganda part (grey) was used for semi-supervised learning (SSL).

and the test set only for evaluation after training. For the experiments, we used two ways of splitting, *data version 1* and *data version 2*. For data version 1, we split Rwanda along its longitude into eleven parts. The left-most and the rightmost five parts were used as the train set, the left-most center part was used as the validation set, and the rightmost center part was used as the test set. In addition, we used three districts in the western part of Uganda as the test set (highlighted area in Figure 5). Those three districts Ibanda, Kasese, & Kabarole are where BBNP completed a trailbridge assessment of the entire district. This data version is more complex, as the test set has in-domain (Rwanda) and out-domain (Uganda) test instances. For data version 2, similar to data version 1, we split Rwanda along its longitude into eleven parts and used ten parts for training and one for validation. In addition to that, we also split Uganda into eleven parts. We refer to the Appendix for a more fine-grained visualization of the data split.

Supervised and semi-supervised model architecture, training and optimization. As the base of our supervised model, we used commonly used convolutional architectures such as ResNet-18, ResNet-50 and WideResNet-50. We trained the supervised model for 200 epochs and used a batch size of 128 and a learning rate of 0.0001 for the Adam optimizer. For testing, we used the model with the highest validation accuracy. In contrast to RGB images used for these network architecture, our data input has more than three channels. We replaced the first convolutional layer with one that matches our data's input channel. We either trained all layers of the network or freeze all layers but the first convolutional and last 9 layers (3 residual blocks). We performed extensive hyperparameter search for our semi-supervised MixMatch model that included model architecture (ResNet-18, ResNet-50 and WideResNet-50), optimization hyperparameter (learning rate, batch size) and model hyperparameters. We selected the model to report based on the best validation accuracy. We refer to the Appendix for a detailed overview of the hyperparameter grid values.

Comparisons We extracted satellite images from the Sentinel-2 MSI: MultiSpectral Instrument, Level-1C (Drusch et al. 2012) for three different tile sizes (320m, 600m, and 1200m). We used a ResNet-50 to train the extracted satellite images. We used a batch size of 128 and a learning rate of 0.0001 for the Adam optimizer. We vary the number of trainable layers, for full (all layers), and transfer (only the last nine layers) learning.

Main Results

We present the balanced accuracy scores of the supervised and semi-supervised models in Table 1. We compared the approaches with different combinations and types of data modalities. We observe that the model trained on only satellite imagery ("only SI") performs poorly irrespectively whether we trained the entire network or just the last nine layers. The semi-supervised model performs best, outperforming all models by at least 10% balanced accuracy. Furthermore, compared to training with only satellite imagery, our semi-supervised model achieved a significant increase in performance, i.e., approx. 34% and 61% boost in performance for data version 1 and 2, respectively. When training only with the best performing data modalities (admin boundaries, OSM base image, OSM waterways, and slope), the ResNet-50 model already achieved high performance. The different data versions showed a significant performance gap between the two datasets. The test dataset that only has in-domain test instances performance approx. 10% better than the one with out-domain test instances (unknown Uganda test instances) even if the semi-supervised approach uses unlabelled Uganda data for training. When using all data modalities, the full ResNet-50 model only performs similarly or slightly better than the one using the best four data modalities. We also plotted the prediction results as an ROC curve in Figure 6. Additional results for tile sizes 300m and 600m, as well as weighted F1 scores for all models per tile size, can be found in the Appendix.

Ablation Study

We perform several experiments to get better insights into the impact of data modalities, data augmentation, averaging inference, pretrained weights and tile sizes.

Impact of Data Modalities and Tile Sizes We investigated which data modalities provided a relatively valuable signal in determining the location of a bridge. We used an identical training infrastructure – including hyperparameters

Model	Data Modalities			Data version 1 *in- and out-domain test*			Data version 2 *in-domain test*		
	w/o SI	AB, OSM-I, OSM-W, SL	Only SI	Rwanda *test*	Uganda *test*	*avg. test*	Rwanda *test*	Uganda *test*	*avg. test*
ResNet-50 (transfer)			✓	49.06 ±0.90	66.53 ±1.77	56.10 ±0.77	49.10 ±0.23	64.11 ±4.97	54.26 ±0.70
ResNet-50 (full)			✓	49.50 ±1.14	69.80 ±0.81	57.69 ±0.75	42.79 ±0.92	63.86 ±2.27	48.85 ±0.81
ResNet-50 (transfer)		✓		65.00 ±1.33	66.07 ±0.47	68.20 ±0.51	62.52 ±0.43	78.91 ±1.56	75.08 ±0.67
ResNet-50 (full)		✓		76.76 ±0.75	66.67 ±0.18	74.57 ±0.24	78.92 ±1.81	78.39 ±1.45	78.49 ±0.51
MixMatch (ResNet-50, full)		✓		81.67 ±0.71	61.20 ±1.59	75.13 ±1.14	82.12 ±0.66	76.82 ±0.94	85.29 ±0.68
ResNet-50 (transfer)	✓			61.80 ±2.67	68.60 ±0.76	68.17 ±0.51	58.65 ±2.64	61.51 ±1.71	74.23 ±1.20
ResNet-50 (full)	✓			80.09 ±0.60	68.67 ±0.93	75.70 ±0.75	70.27 ±4.73	83.98 ±1.17	81.99 ±1.17
MixMatch (ResNet-50, full)	✓			**82.97** ±0.84	**69.20** ±0.87	**77.69** ±0.83	**82.09** ±0.34	**84.77** ±2.73	**87.79** ±0.69

Table 1: **Bridge site estimation (1200m) balanced accuracy results (higher is better)**. We present results (mean and standard error for three runs each) for different models and different modalities. We used ResNet-50 as the architecture backbone, and trained the entire network (*full*) or used it as transfer learning in which only the last N layers were trained (*transfer*). We varied the types of data modalities used for training. We had three data modality settings: 1) all types of geospatial data except for satellite imagery (*w/o SI*), 2) the best four performing data modalities admin boundaries (*AB*), OSM base image (*OSM-I*), OSM waterways (*OSM-w*) and slope (*SL*), and 3) only satellite imagery (*Only SI*). We report balanced accuracy for both data versions 1 and 2 (with a detailed country-wise split) as mean and standard error over three independent runs with different seed.

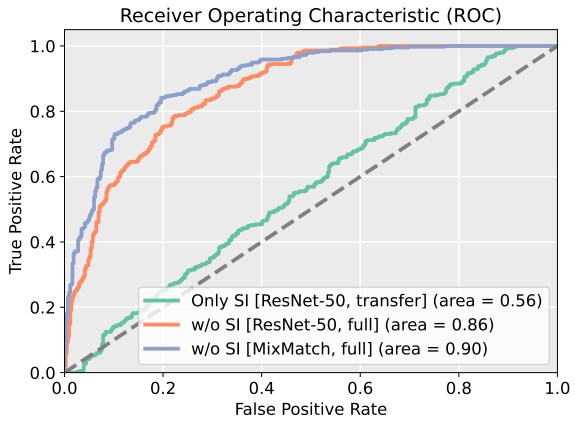


Figure 6: **Receiver Operating Characteristic (ROC) for data version 2.** We plot the ROC of the baseline with satellite imagery ("Only SI") against our supervised approach (w/o SI [ResNet50, full]) and our semi-supervised approach (w/o SI [MixMatch, full]).

and model architecture – to train each data modality alone. We refer to the Appendix for the experimental setup. Our findings are visualized in Figure 7 in which we plot the validation accuracy during training for each tile size. For all tile sizes, waterways, administrative bounds, slope, and OSM image provide the most valuable and consistently reliable signal regardless of the input size, resulting in high validation accuracies. The Population and OSM Roads modalities do not - by themselves - provide a valuable signal. The validation accuracy of the elevation data modality is highly variant throughout the training run for each tile size – likely indicating overfitting that is corrected as training continues. The training on separate data modalities also showed no significant impact of tile sizes on the model performance. We observed with further experiments in the Appendix that an

Model arch.	Augmentation	Average Inference	Pretrained Weights	Test Acc.
ResNet-18	✓	✓	✓	83.24 ±0.40
ResNet-50	✗	✗	✗	77.92 ±0.83
ResNet-50	✓	✗	✗	79.22 ±0.50
ResNet-50	✗	✓	✗	80.26 ±0.41
ResNet-50	✗	✗	✓	79.19 ±0.97
ResNet-50	✓	✓	✓	81.57 ±0.35
WideResNet-50	✓	✓	✓	79.80 ±0.85

Table 2: **Impact of augmentation, average inference and pretrained weights.** We measure the impact of data augmentation, average inference and pretrained weights with the balanced test accuracy of data version 2. The test accuracies are reported as mean and standard error over three runs.

increase in tile size only led to minor performance gains.

Impact of Data Augmentation, Averaging Inference and Pretrained Weights We evaluated the impact of different design choices (data augmentation, averaging inference, and pretrained weights) by training supervised deep learning models with the same hyperparameter and optimization settings while varying one or several of these choices. The results of this ablation are shown in Table 2. The experiments showed that all three features lead to a performance gain, with the combination of all three choices leading to the best performance for ResNet-50 (77.92% vs. 81.57% balanced accuracy). Surprisingly, we achieved better overall performance with ResNet-18 than with ResNet-50 or WideResNet-50. We hypothesize that smaller network architectures are more suitable and less prone to overfitting due to the small dataset size. However, we have observed consistently better performance with ResNet-50 for our semi-supervised experiments.

Discussion

Is every river crossing a potential site? BBNP considers every river crossing in which the local community states a trailbridge is needed to be a potential site. The only sites considered not to be potential sites are where the community

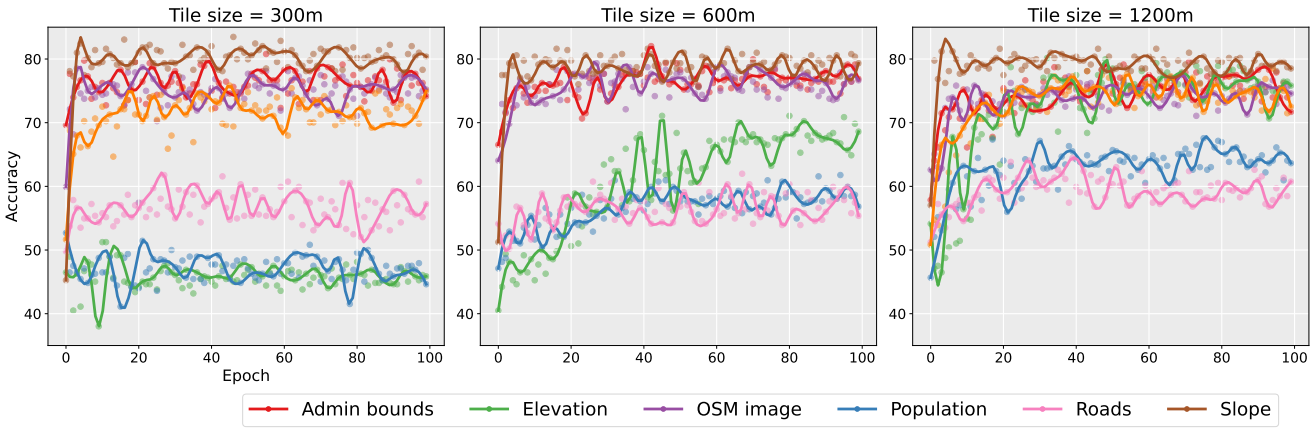


Figure 7: **Impact of single modalities**: Validation loss during training with identical model architecture and optimization for different data modalities.

only wants a vehicle bridge or an existing bridge present or within 300 meters that provides the community with year-round access to cross the river. If many river crossings were identified in adjacent tiles along the same waterway, BBNP would manually analyze the best location for a trailbridge. Local communities ultimately determine whether a trailbridge is needed at a site, and governments ultimately decide how transportation budgets are allocated. BBNP intends to use the remotely identified river crossings as an estimate for the overall need in a region and starting point for where to conduct field assessments rather than to determine where bridges will ultimately be constructed.

Why did we not assume that the bridge site is in the center of the river? The GPS coordinates which are used for training data are collected when there is usually no bridge present at the river crossing, thus the reason the community is requesting a bridge. Therefore, the assessors would stand at the river's edge when collecting the coordinates. Putting the coordinates precisely in the middle of the river would require manually checking and modifying every training point.

How likely are "negative" tiles negative? How likely is it that BBNP missed identifying all bridge sites in Rwanda We expect to miss some potential sites when conducting a full coverage assessment, i.e., the Rwanda nationwide needs assessment and the full coverage assessments in the three districts of Uganda. In order to minimize the number of sites missed, BBNP followed up with all local administrations one level above the village level that submitted zero requests to confirm there was no need for a trailbridge in their domain. Additionally, assessors inquired about additional sites at every site they visited. Currently, BBNP has 1,438 locations where trailbridges have been requested by communities in Rwanda. Since the 2018 Rwanda needs assessment, 21 of the 1,438 sites were submitted after the assessment was concluded. Over 100 other sites have been submitted to BBNP in the last five years and were found to have already been located and assessed. Despite the trailbridge program becoming more popular and known throughout Rwanda and the Ugandan districts, fewer additional requests come in each year, with a lower percentage of sites being new requests each year. Building Bridges Non-Profit is confident they have identified over 90 percent of the potential trailbridge sites in Rwanda and the three districts assessed in Uganda.

Conclusion & Future Work

In this work, we presented an approach for incorporating multi-modal geospatial data for identifying river crossing for trailbridges. We showed that using satellite imagery results in poor performance and that using multi-modal non-satellite-imagery data leads to a significant increase in performance (at least 34% increase in balanced accuracy). This initial investigation showed immense potential for using multi-modal geospatial data. Future work will explore additional ways to leverage more unlabelled data (e.g., semi-supervision or self-supervision) and to fuse multi-modal geospatial data (e.g., early or late fusion) to improve predictive performance. Furthermore, combining high-resolution imagery and our chosen data modalities is a potential avenue of investigation.

One of the main purposes of this project was to determine which data modality is most effective in locating potential trailbridge sites, i.e., river crossings, and how effective these data modalities could be compared to satellite imagery. Waterways, admin boundaries, OSM base image, and slope were found to be the most effective. Finding a substitute for insufficient waterways datasets was one of the analysis's most critical and challenging aspects. Future work will explore incorporating more prior knowledge. For instance, sections of an administrative boundary that are not sinuous are unlikely to be a waterway, such as sections or slope values perpendicular to the line can eliminate the possibility of water, e.g., a ridgeline or hillside.

Broader Impact Recent research has extensively focused on deep learning approaches for satellite imagery. We showed that using different types of more processed geospatial data can be highly beneficial for predictive task perfor-

mance. We hope our investigation (extraction, data augmentation and training) will provide insights on using these data modalities effectively for deep learning approaches.

The ability to quickly and cost-effectively estimate trail-bridge needs across a region or nation is key to BBNP's strategic priorities. It will support productive conversations with governments and other influential stakeholders at the national level. Further, it will equip BBNP with information to advise governments and other influential stakeholders on long-term rural transportation infrastructure plans. By scaling and automating the efforts of remote site assessment, we provide funders of rural infrastructure impact estimates for household income, reduction in travel time, market access, school enrollment, truancy, and access to health centers. In doing so, we promote increased investment in rural transportation infrastructure and emphasize its importance in eliminating poverty on a global scale.

Acknowledgments

We would like to thank the Autodesk Foundation for their generous support of BBNP's work, including the opportunity to collaborate with experts on important technical topics such as those described in this work. We would also like to thank the governments of Rwanda and Uganda for their dedicated partnership in creating reliable rural access.

We would like to thank Stefano Toparini and Colm Ferrari for helping with the first project involving investigating waterway datasets, and bridge type and span classification. We would like to thank Jeffrey Landes for his initial model evaluation.

References

Audebert, N.; Le Saux, B.; and Lefèvre, S. 2017. Joint learning from earth observation and openstreetmap data to get faster better semantic maps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 67–75.

Audebert, N.; Le Saux, B.; and Lefèvre, S. 2018. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS journal of photogrammetry and remote sensing*, 140: 20–32.

Audebert, N.; Saux, B. L.; and Lefèvre, S. 2016. Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. In *Asian conference on computer vision*, 180–196. Springer.

Bennett, J. 2010. *OpenStreetMap*. Packt Publishing Ltd.

Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; and Raffel, C. A. 2019. Mixmatch: A holistic approach to semi-supervised learning. *Advances in Neural Information Processing Systems*, 32.

Brooks, W.; and Donovan, K. 2021. Eliminating Uncertainty in Market Access: The Impact of New Bridges in Rural Nicaragua. *Econometrica*, 88(145275).

Brown, M.; Goldberg, H.; Foster, K.; Leichtman, A.; Wang, S.; Hagstrom, S.; Bosch, M.; and Almes, S. 2018. Large-scale public lidar and satellite image data set for urban semantic labeling. In *Laser Radar Technology and Applications XXIII*, volume 10636, 154–167. SPIE.

Crippen, R.; Buckley, S.; Belz, E.; Gurrola, E.; Hensley, S.; Kobrick, M.; Lavalle, M.; Martin, J.; Neumann, M.; Nguyen, Q.; et al. 2016. NASADEM global elevation model: methods and progress.

Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment*, 120: 25–36.

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; and Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202: 18–27.

Hassani, H.; Huang, X.; and Silva, E. 2019. Big data and climate change. *Big Data and Cognitive Computing*, 3(1): 12.

Hu, J.; Mou, L.; Schmitt, A.; and Zhu, X. X. 2017. FusioNet: A two-stream convolutional neural network for urban scene classification using PolSAR and hyperspectral data. In *2017 Joint Urban Remote Sensing Event (JURSE)*, 1–4. IEEE.

Kawasaki, A.; Berman, M. L.; and Guan, W. 2013. The growing role of web-based geospatial technology in disaster response and support. *Disasters*, 37(2): 201–221.

Kiwelekar, A. W.; Mahamunkar, G. S.; Netak, L. D.; and Nikam, V. B. 2020. Deep Learning Techniques for Geospatial Data Analysis. In *Machine Learning Paradigms*, 63–81. Springer.

Konisky, D. M.; Hughes, L.; and Kaylor, C. H. 2016. Extreme weather events and climate change concern. *Climatic change*, 134(4): 533–547.

Long, Y.; and Shen, Z. 2015. *Geospatial analysis to support urban planning in Beijing*. Springer.

Manfré, L. A.; Hirata, E.; Silva, J. B.; Shinohara, E. J.; Giannotti, M. A.; Larocca, A. P. C.; and Quintanilha, J. A. 2012. An analysis of geospatial technologies for risk and natural disaster management. *ISPRS International Journal of Geo-Information*, 1(2): 166–185.

Nash, E.; Korduan, P.; and Bill, R. 2009. Applications of open geospatial web services in precision agriculture: a review. *Precision agriculture*, 10(6): 546–560.

Nations, T. U. 2019a. Infrastructure: Underpinning Sustainable Development. https://unops.economist.com/wp-content/uploads/2019/01/Infrastructure_underpining_sustainable_development_EN.pdf.

Nations, T. U. 2019b. The Sustainable Development Report: 2019. https://unstats.un.org/sdgs/report/2019/The-Sustainable-Development-Goals-Report-2019.pdf.

Ojogbane, S. S.; Mansor, S.; Kalantar, B.; Khuzaimah, Z. B.; Shafri, H. Z. M.; and Ueda, N. 2021. Automated Building Detection from Airborne LiDAR and Very High-Resolution Aerial Imagery with Deep Neural Network. *Remote Sensing*, 13(23): 4803.

Praveen, B.; and Sharma, P. 2020. A review: The role of geospatial technology in precision agriculture. *Journal of Public Affairs*, 20(1): e1968.

Roberts, P.; KC, S.; and Rastogi, C. 2006. Rural Access Index: A Key Development Indicator. *The World Bank Group: Transport Papers*, TP-10.

Sun, Y.; and Du, Y. 2017. Big data and sustainable cities: applications of new and emerging forms of geospatial data in urban studies.

Tarvainen, A.; and Valpola, H. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30.

Tatem, A. J. 2017. WorldPop, open data for spatial demography. *Scientific data*, 4(1): 1–4.

Thomas, E.; Bradshaw, A.; Mugabo, L.; MacDonald, L.; Brooks, W.; Dickinson, K.; and Donovan, K. 2020. Engineering environmental resilience: A matched cohort study of the community benefits of trailbridges in Rural Rwanda. *Science of the Total Environment*, 771(5): 1965–1997.

Zhang, H.; Cisse, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2018. mixup: Beyond Empirical Risk Minimization. In *International Conference on Learning Representations*.

Finding and Building Bridges: Details

Remote assessment and criterion BBNP's strategy for addressing rural access at scale includes supporting national governments in incorporating pedestrian and motorcycle infrastructure into their transportation planning. To support that work, BBNP conducts needs assessments to determine the scope and distribution of trailbridge needs. Specifically, the target is to locate rivers or other barriers, e.g., ravines

1. where pedestrians, bicyclists, livestock, and light motorized traffic need to cross to reach important destinations, e.g., markets, farms, healthcare centers, and schools, and
2. where an all-weather crossing, e.g., a vehicular bridge that is passable year-round, doesn't exist within 300 meters.

Trailbridges The bridges are designed to handle live loads such as pedestrians, motorbikes, and livestock such as cattle. The bridges' decks are intentionally narrow (1 meter-wide) so that heavier vehicles can not attempt to pass. Depending on river characteristics, terrain, and the span, trailbridges can utilize various designs such as steel truss, concrete, stone arch, concrete, wooden, and cable suspended. BBNP focuses primarily on cable suspended bridges, as these are the most efficient for spanning longer distances (30 meters or more) where local governments typically request assistance with designing, funding, and construction.

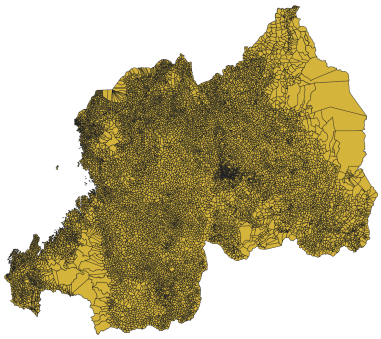
Identifying trailbridge sites In 2018, BBNP conducted a nationwide needs assessment in Rwanda. They contacted local leaders at the sector or cell level across the entire country, informed them about the trailbridge needs assessment, and provided the opportunity to submit trailbridge requests for their communities. BBNP's call center would follow up with local officials for any cells and sectors that did not submit trailbridge requests to confirm they knew about the assessment and did not need any trailbridges. Over 1,500 sites were submitted using this process, and after vetting the sites for invalid submissions such as vehicular bridges, sites were visited for in-field assessments. While conducting field assessments, assessors would ask if any other crossings needed a trailbridge in the area. Assessors were assigned to specific districts and tasked with confirming that all sites had been identified in a sector before moving to the next. The same process was utilized in Uganda for three districts, and these sites, combined with the Rwanda sites are the training data for remote site identification. While the nationwide needs assessment's original purpose was not to produce training data, its objective was to identify every river crossing needing a trailbridge in all of Rwanda.

Whether or not a river crossing is a potential site comes down to the local community's opinion and a few other factors. The most common reason for a river crossing not to be a potential trailbridge is because the community *only* wants a vehicular bridge. In rural areas of East Africa, particularly Uganda, it is common for a path or road to be wide enough for vehicles to pass while the main form of transportation is walking and motorbikes. For this reason, BBNP still considers these crossings as potential sites, and it is common for communities to prefer a vehicle bridge but still accept a trailbridge. Ultimately, the local district/county government decides on which type of bridge to invest in, as a vehicle bridge is many times more expensive. Often at these locations, vehicles can pass during the dry season, and trailbridges are cost-efficient solutions for a 3-4 month rainy season. A notable exception to a river crossing not being a potential site is when a path crossing a waterway is present due to few individuals crossing or animals/livestock. these paths are sometimes referred to as "game trails" or "livestock trails". Trailbridges are typically placed along the main route between villages. These crossings can be determined when river crossings are far from communities without any sign of a reasonable path from likely origins and destinations. Therefore, we assume that if there is a path and a barrier, e.g., river, and no bridge is present, than that location needs a bridge. It is also common for communities to request a bridge at river crossings where they have a log bridge when the log bridge is destroyed each year after the first heavy rainfall. The goal for this project was to identify all likely pedestrian river crossings and then manually determine if a bridge is needed until we have more data to train the model on differentiating itself.

Data Extraction, Usage and License

We extracted the geospatial data from public resources. In what follows, we explain the choice of data modalities before detailing the process of data extraction and preprocessing. Finally, we briefly state the data licence for each data modality.

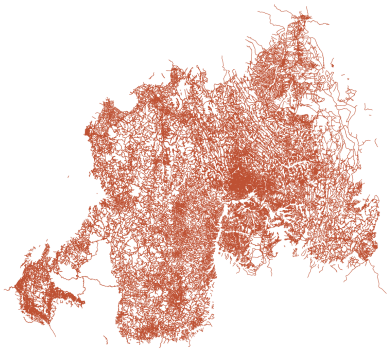
Choice of Data Modalities Most of Building Bridges Non-Profit's existing trailbridges in Rwanda and Uganda are along rivers that are not represented in the countries' waterways data. BBNP has noted that most constructed bridges connected two local administrations, i,e, sector, parish, or county. This is not surprising, as rivers make natural landmarks for administrative boundaries. Therefore, administrative boundaries were included as a potential alternative or supplement to insufficient waterways data. Terrain elevation, slope, OSM roads, and waterways were included to indicate whether a river or gorge is present in the tile. We investigated whether raster population data (100-meter resolution) would contribute to the model by informing it where people live, including the more densely populated areas where common destinations of travelers – markets, schools, and health centers – are likely to be. Vast areas without population e.g. protected areas, game reserves, large farms, etc. that do not need bridges due to the lack of pedestrian traffic are also visible in this dataset. In addition to extracting specific layers of the OSM, we also extracted the entire map as an image. First, standard deep learning models have achieved outstanding performances for images, and thus, using images could ease the transfer of pre-trained models to the task of remote site identification. Second, when extracting roads and waterways from OSM, we noticed that many more minor roads and waterways were not "labele" as such. However, they are visible in the "image" maps of OSM.



(a) Admin boundaries

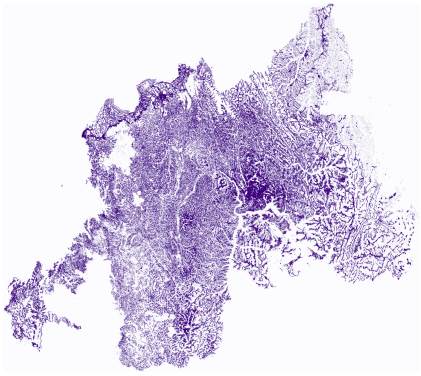
(b) Elevation

(c) OSM Image

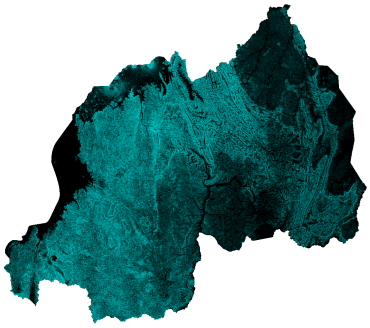


(d) OSM Roads

(e) OSM Waterways



(f) Population



(g) Slope

Figure 8: **Visualization of Rwanda geospatial data used for extraction.** These data modalities are rasterized, augmented and used for training.

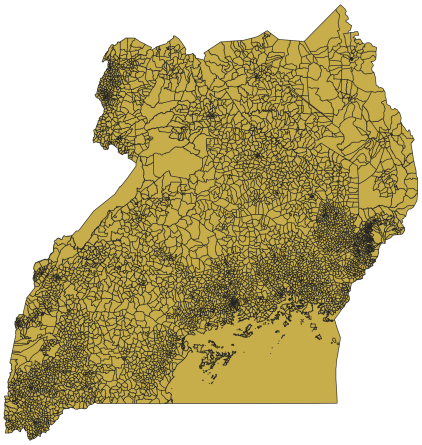
Detailed Data Extraction and Pre-Processing

If not stated, we rasterize the data with an output resolution of 1/3600, i.e., 1-arc.

Administrative Boundaries The admin boundary datasets consists of shapefiles of polygon geometries. If we rasterize these shape file, our rasterization will produce large polygons, instead of, what we want, lines that represent the boundaries. Therefore, we first convert the polygons to line string geometries. After the conversion, we rasterize the shapefile. The raster contains only two unique pixel values – value 1 represents the presence of an administrative boundary, and value 0 represents its absence.

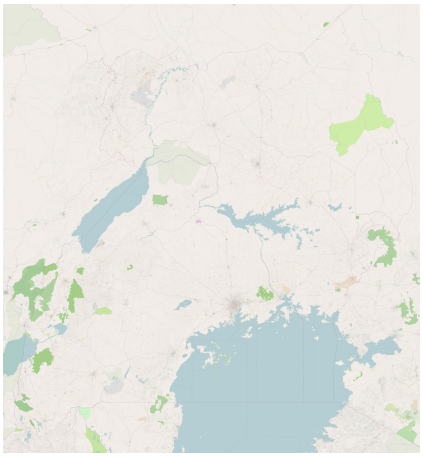
Elevation and Slope We used the Google Earth Engine[4] (Gorelick et al. 2017) to extract elevation and slope rasters from the NASADEM digital elevation model. We constraint the elevation values to $[0, 9000]$ and slope values

[4]https://earthengine.google.com/

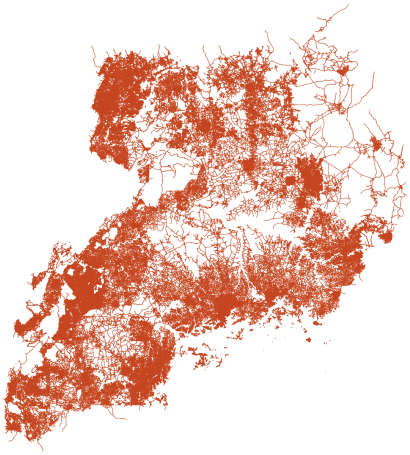


(a) Admin boundaries

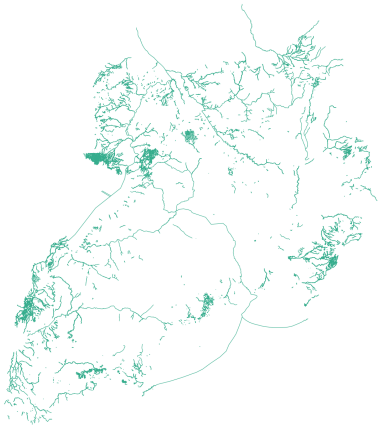
(b) Elevation



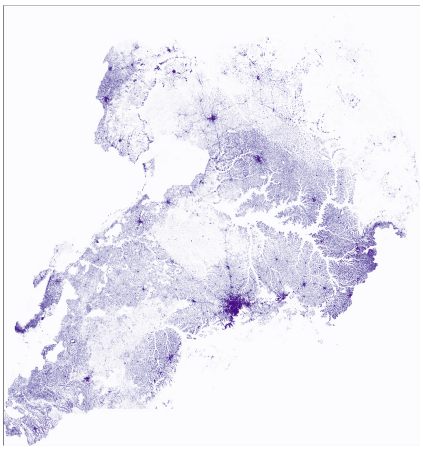
(c) OSM Image



(d) OSM Roads



(e) OSM Waterways



(f) Population

(g) Slope

Figure 9: **Visualization of Uganda geospatial data used for extraction.** These data modalities are rasterized, augmented and used for training.

to $[0, 90]$. We scale and quantize the values to an image value space $[0, 255]$ and saved the image to the red channel of the GeoTIFF file. The resolution of the dataset is 30m per pixel.

Model	Data Modalities			Data version 1 *in- and out-domain test*			Data version 2 *in-domain test*		
	w/o SI	AB, OSM-I, OSM-W, SL	Only SI	Rwanda *test*	Uganda *test*	*avg. test*	Rwanda *test*	Uganda *test*	*avg. test*
ResNet-50 (transfer)			✓	0.49 ±0.01	0.66 ±0.01	0.56 ±0.01	0.45 ±0.02	0.64 ±0.02	0.52 ± < 0.01
ResNet-50 (full)			✓	0.47 ±0.01	0.69 ±0.01	0.56 ±0.01	0.41 ±0.01	0.64 ±0.03	0.48 ±0.01
ResNet-50 (transfer)		✓		0.61 ±0.01	0.64 ±0.01	0.68 ±0.01	0.58 ±0.01	0.82 ±0.02	0.75 ±0.01
ResNet-50 (full)		✓		0.78 ± < 0.01	0.63 ±0.01	0.74 ± < 0.01	0.79 ±0.01	0.77 ±0.02	0.83 ±0.01
MixMatch (ResNet-50, full)		✓		**0.82** ±0.01	0.55 ±0.03	0.75 ±0.01	0.82 ±0.01	0.76 ±0.01	0.85 ±0.01
ResNet-50 (transfer)	✓			0.58 ±0.02	0.66 ± < 0.01	0.68 ±0.01	0.53 ±0.03	0.81 ±0.01	0.74 ±0.01
ResNet-50 (full)	✓			0.78 ±0.01	0.66 ±0.01	0.75 ±0.01	0.72 ±0.03	**0.83** ±0.01	0.82 ±0.01
MixMatch (ResNet-50, full)	✓			**0.82** ± < 0.01	**0.68** ±0.03	**0.78** ±0.01	**0.83** ±0.01	0.82 ±0.03	**0.87** ±0.01

Table 3: **Bridge site estimation (1200m) weighted F1 score results (higher is better)**. We present results (mean and standard error for three runs each) for different models and different modalities. We used ResNet-50 as the architecture backbone, and trained the entire network (*full*) or used it as transfer learning in which only the last N layers were trained (*transfer*). We varied the types of data modalities used for training. We had three data modality settings: 1) all types of geospatial data except for satellite imagery (*w/o SI*), 2) the best four performing data modalities admin boundaries (*AB*), OSM base image (*OSM-I*), OSM waterways (*OSM-w*) and slope (*SL*), and 3) only satellite imagery (*Only SI*). We report balanced accuracy for both data versions 1 and 2 (with a detailed country-wise split) as mean and standard error over three independent runs with different seed.

Tile size	Type	Data version 1	Data version 2
300m	supervised	72.10 ±0.63	83.45 ±0.95
600m	supervised	75.70 ±0.26	81.26 ±1.03
1200m	supervised	75.13 ±1.14	85.29 ±0.68

Table 4: **Impact of tile size**. Balanced accuracy (higher is better) of the best performing approaches w.r.t. tile size.

Population The population data are provided as rasterized GeoTIFF files. The data resolution is 100m.

OpenStreetMap OSM provides a host of vector-based geometries around - from roads to buildings to parks. We identified two potentially valuable subsets of this data - waterways and roads. We first performed discretization by mapping categorical values, e.g., the type of road or waterway, to a scalar value that can be used in rasterization. After rasterization, we ignore all subtypes of roads and waterways and set all road and waterway pixels to 1, all non road and water way pixels to 0. In addition to extracting specific layers of the OSM, we also extracted the entire map as an image.

After rasterization, we used both Rwanda and Uganda data to calculate the mean and standard deviation of all data channels to normalize the data for training and evaluation. We have visualized all data modalities for Rwanda and Uganda in Figures 8 and 9, respectively.

Data Licence

The Rwanda administrative boundaries data is publicly available through the National Institute of Statistics of Rwanda and can be found here[5]. The Uganda admin boundaries were extracted from CloudMade data, derived from OpenStreetMap (www.openstreetmap.org) and is made available by MapCruzin (www.mapcruzin.com) and is licensed under the terms of the Creative Commons Attribution Share-Alike 2.0 license. Earth Engine is a platform for scientific analysis and visualization of geospatial datasets and can be used for academic and non-profit purposes. The WorldPop population data is published under the Creative Commons Attribution 4.0 International License[6], i.e., users are free to use as long as clear attribution of the source is provided. OSM data is free to use for any purpose.

[5]https://www.statistics.gov.rw/publications/rwanda-administrative-map

Additional Evaluation Details and Results

Data Versions 1 and 2. For data version 1, we have 815 training bridge sites (all in Rwanda), 242 validation bridge sites (all in Rwanda) and 620 test bridge sites (370 in Rwanda, 250 in Uganda). For data version 2, we have 982 bridge sites (815 in Rwanda, 167 in Uganda), 261 validation bridge sites (242 in Rwanda, 19 bridge sites in Uganda) and 434 test bridge sites (370 in Rwanda, 64 in Uganda). We sampled as many negative sites as there are positive sites.

Data V1 splits Rwanda into 10 parts equally by its longitude. Parts 1-3 and 7-10 are used for train, part 4 is used for validation, parts 5, 6, and Uganda are used for testing. Data V2 uses Rwanda as in Data V1 and further splits Uganda into 10 parts equally by its longitude. Parts 1-2 are used for testing and 3-8 are used for training, parts 9-10 are used for validation. We visualized the split in Figure 10.

Hyperparameters for the Semi-Supervised Evaluation with MixMatch We evaluated on a large grid of hyperparameter for MixMatch and chose the model with the best validation performance. We used the following hyperparameter space:

- $\lambda_{\mathcal{U}} = [[0, 0.01, 0.1, 1.0]$
- EMA decay rate $= [0.5, 0.75, 0.9]$
- $T = [0.5, 0.75]$
- $\alpha = [0.25, 0.5, 0.75]$

Implementation and Open Source

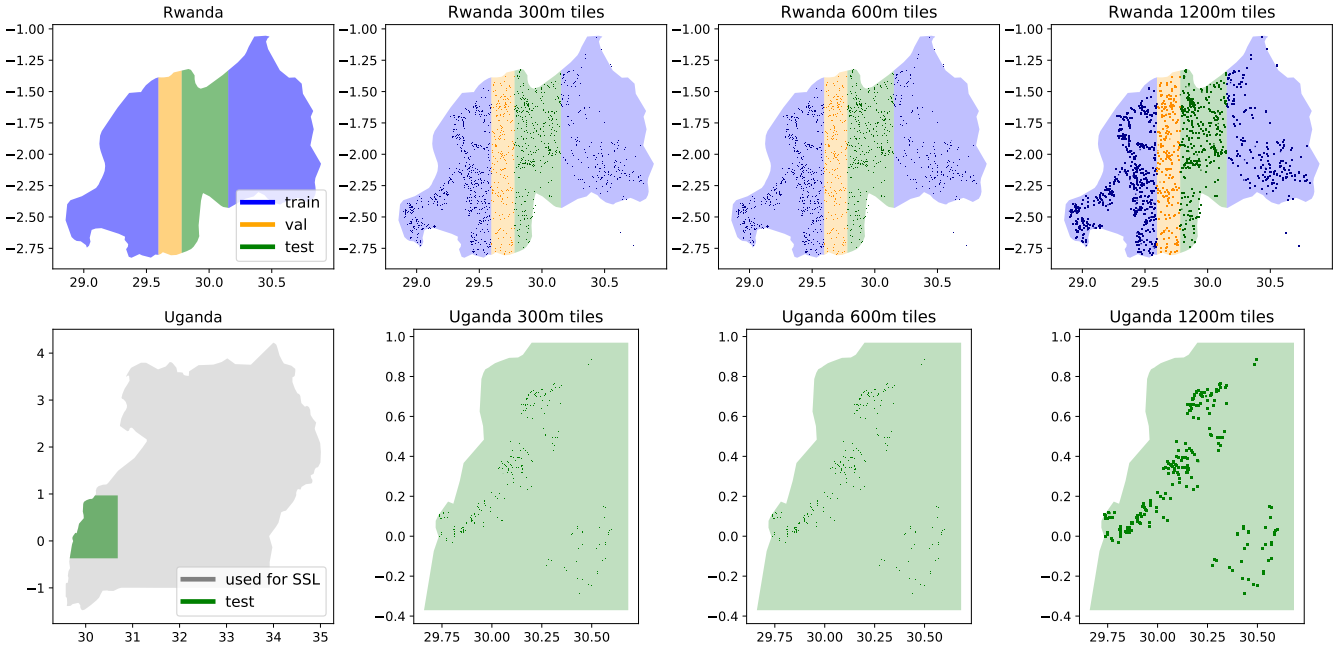
We implemented the entire work in PyTorch and have made it available in the Supplementary Materials. Upon accep-

[6]http://creativecommons.org/licenses/by/4.0

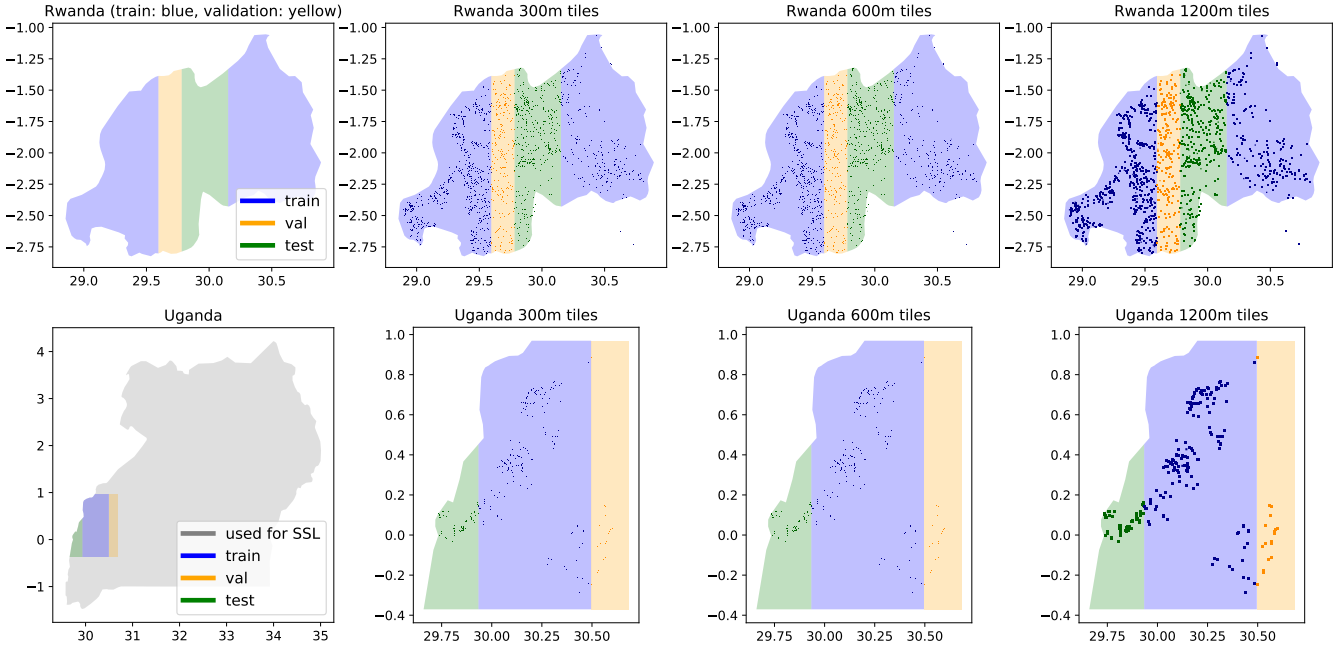
tance, both code implementation and data will be open-sourced for research-only usage.

Additional Evaluation Results

The balanced F1 scores for 1200m tiles are found in Table 3.



(a) Data version 1



(b) Data version 2

Figure 10: **Visualization of train, validation, test split of data version 1 and 2.** We split the labelled data according to its geographic location into different sets (train, validation and test). The unlabelled Uganda part (grey) was used for semi-supervised learning (SSL). The scattered points on the subplots represent the trailbridge locations.