Analysis of Deep Learning Techniques for Maasai Boma Mapping in Tanzania

Keli Cheng, *Student Member, IEEE*, Ilinca Popescu, Lincoln Sheets, and Grant J. Scott, *Senior Member, IEEE*

***Abstract*—Underdeveloped countries in sub-Saharan Africa often contain cultural subpopulations that are underserved in regard to health and education. This perpetuates the health challenges of the country as a whole, and it is therefore of interest to be able to automatically map the subpopulation for the health services delivery. International nonprofit health organizations have often taken the lead in these efforts, providing humanitarian aid (e.g., clean water and food) as well as health care. This is necessary, as the ethnic subpopulations are not well integrated into the society and the existing health care systems. In this study, we explicitly explore the Maasailand of Tanzania, to evaluate the use of deep neural networks (DNN) to aid in the automatic visual analysis of remote sensing data to geolocate Maasai boma structures. We investigate the performance of four state-of-the-art DNN as classifiers of boma presence within high-resolution imagery; all showing over 95% F1 score performance. Additionally, we scan over 3900 km^2 of high-resolution imagery, combining a ProxylessNAS with broad area aggregation and mapping techniques and demonstrate the discovery of hundreds of boma, many that were not discovered by human analysts performing visual scans. The trained ProxylessNAS model generates a classified vector response field (CVRF). The CVRF is aggregated by a mode-seeking algorithm to detect potential locations of boma structures within the study area. The model detected numerous human false negatives (HFNs) and achieved 94.022% TPR and 95.395% F1 score using an aggregation aperture of 250 m within a 76.620 square kilometers area of interest.**

***Index Terms*—Boma, broad area search (BAS), deep learning.**

I. Introduction

SUB-SAHARAN Africa includes a diverse set of underdeveloped countries that struggle with health matters related to basic human needs. Women and children are especially vulnerable in light of the potential health complications from unclean water and insufficient medical care, where access to modern health care and sanitation technologies are not readily available. "Globally, 50% of children under five who die of pneumonia, diarrhea, measles, HIV, tuberculosis, and malaria are in Africa, according to the World Health Organisation (WHO)." [1] Numerous international nonprofit groups provide targeted support for these populations through a variety of food aid and health clinic services efforts. Medical services, food distribution, and clean water initiatives are critical to reducing the mortality of children and women in sub-Saharan Africa, and *Humanity for Children* (HFC) is one such international, nonprofit organization that seeks to address the health needs for the Maasailand in northern Tanzania. HFC has initiated a program to map the remote villages in Maasailand of Tanzania for the purpose of providing mobile health care clinics, called *Community Health Assessment Mapping Project* (CHAMP).

The Maasai people is an ethnic population in Tanzania that are not well integrated into the country's existing public health services. The Maasai villages, called bomas, have populations ranging from ten to over 200 inhabitants. HFC is developing mobile health clinics to serve the needs of expectant mothers and children throughout the distributed bomas of Maasailand. Since the location of bomas are not documented within existing geographic information systems (GIS), it is time consuming but necessary to physically travel and map bomas using ground-based vehicles (e.g., motorcycles). These ground surveys map the locations and paths to bomas by GPS tagging from motorcycles, which is an expensive and time-consuming process. Computational methods, especially contemporary deep learning computer vision models, can be leveraged to aid the task of boma location discovery and mapping. Deep neural networks (DNN) can be applied across a broad swath of area, scanning satellite imagery to provide a very cost-effective and time-saving approach for boma mapping. In [2], it was demonstrated that numerous DNN could achieve high F1 scores, over 95%, in cross-validation experiments of boma recognition. However, it is also necessary to evaluate the use of DNN in real-world applications, outside of the close-world cross-validation paradigm. In this case, we seek to quantify how well DNN can discover unknown boma within the Maasailand.

The use of remote sensing imagery for mapping has traditionally focused on land cover [3], modern infrastructure [4], land forms [5], and crop monitoring [6]. Recently, DNN such as ProxylessNAS [7], Xception [8], Inception-ResNet-V2 [9], and ResNet50 [10] have been applied in a variety of remote sensing use-cases. For example, [11] and [12] have explored scene classification of remote sensing imagery (RSI). Land cover mapping with DNN has also been explored, e.g., [13] and [14]. DNNs have also been used for complex object analysis of RSI, e.g., [15] and [16]. These are just a few of the litany of prior research using DNN on RSI for computer vision and geo-spatial analysis, and this is the inspiration to apply DNN, such as listed above, to provide humanitarian mapping of Maasai bomas in Tanzania.

Manuscript received December 31, 2021; revised March 10, 2022; accepted March 30, 2022. Date of publication April 13, 2022; date of current version May 25, 2022. *(Corresponding author: Grant J. Scott.)*

The authors are with the University of Missouri, Columbia, MO 65211 USA (e-mail: kchkr@mail.missouri.edu; ipopescu@stanford.edu; sheetslr@health.missouri.edu; GrantScott@missouri.edu).

Digital Object Identifier 10.1109/JSTARS.2022.3167373

This article provides a recap from [2] for scene classification performance of various DNN for identifying RSI chips with boma. We then further develop the process to leverage the top-performing DNN for real-world mapping over a large swath of earth in Tanzania. We demonstrate a method that is drastically more efficient than terrestrial visits for mapping with GPS receivers, as is currently done; as well as less error-prone (lower missed detections) than a human visual scan. The final results of our detections are shared with our nonprofit partners to aid their efforts in providing health services in Maasailand of Tanzania.

Image chips covering the study area are acquired and inferred through the trained DNN model to generate a binary classification vector response field (CVRF). A response surface is extracted from the CVRF and a mode-seeking clustering algorithm is applied to estimate the positions of bomas. Furthermore, to provide a quantitative assessment of the suitability of DNN for processing large portions of the earth to triage detections, the scanning precision metric is presented and contrasted with close-world metrics such as traditional F1, precision, and recall. Our results highlight the challenges associated with broad area analytics, as the trained models discover bomas that are missed by human scanning and labeling.

The rest of this article is organized as follows. Section II details our datasets, including imagery and human-mapped (partial ground-truth) location data. Section III provides a summary of the DNN applied for this research and how those DNN can be leveraged in a real-world analytical workload. Experimental methods and results are discussed in Section IV for cross-validation and scanning precision. In Section V, we discuss the key findings and insights. Finally, Section VI concludes this article.

II. Tanzania Maasai Boma Dataset

In [2], the manual curation of the boma dataset using Google Earth and the Google Maps API is detailed. Fig. 1 is an example from Google Earth, with three bomas marked by pushpins. The distinctive visual elements include animal fencing (for herd animals) and small structures for the boma population. As can be observed, the majority of the boma is the same earthen material as the surrounding context, with sporadic structural elements. This is particularly challenging for current DNN techniques that have been shown to be overly reliant on texture [17], where in this case the boma is built in the environment with minimal ground disturbance. Our study area is approximately 3906 km^2.

Our data collection, as detailed in [2], is summarized as follows. Google Earth Pro was used to manually scan the study area, bounded by 3° 58' 37" S, 36° 09' 05" E to 4° 31' 18" S, 36° 43' 53" E, and bomas were marked with dropped pins in the approximated center. The final count of human discovered bomas was 635, afterwhich the coordinates were used to acquire a nominal 0.6 m ground sample distance (GSD) image chips from the Google Maps Static API. Negative examples, i.e., landscape nearby without boma content, were acquired by randomly pulling down image chips in 1 km standoff from boma

Fig. 1. Maasai boma in Tanzania. Imagery copyright 2021, CNES/Airbus.

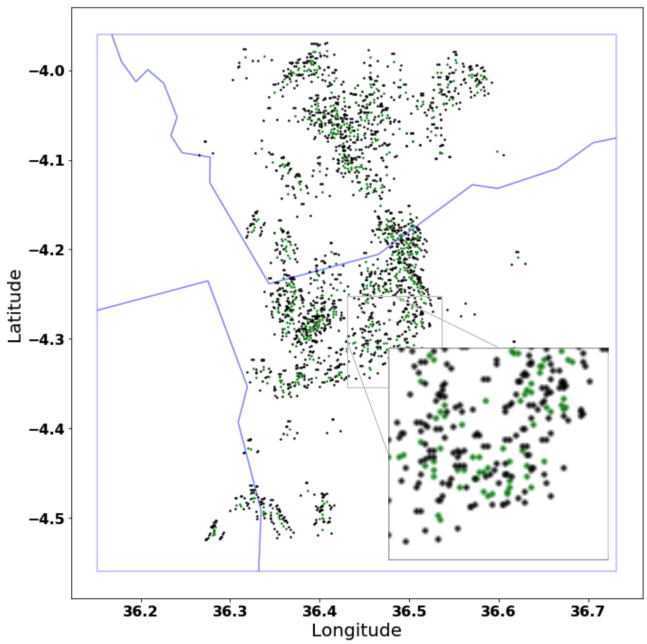


Fig. 2. Geospatial distribution of boma (green) and non-boma (black) within the study area of Tanzania [2].

in various selections of the cardinal and intercardinal directions: North, South, East, West, Northeast, Northwest, Southeast, Southwest. These negative examples were then pruned to remove chips from the negative chips with boma content, resulting in 1726 non-boma training chips. Fig. 2 shows the spatial distribution of the positive (green) and negative (black) boma chips. Fig. 3 provides example of RSI training chips, both boma and non-boma. Though the differences between boma and non-boma image chips are significant, there still exist some irregular circular structures, fencing, and trees in non-boma examples that may mislead the machine during detection.

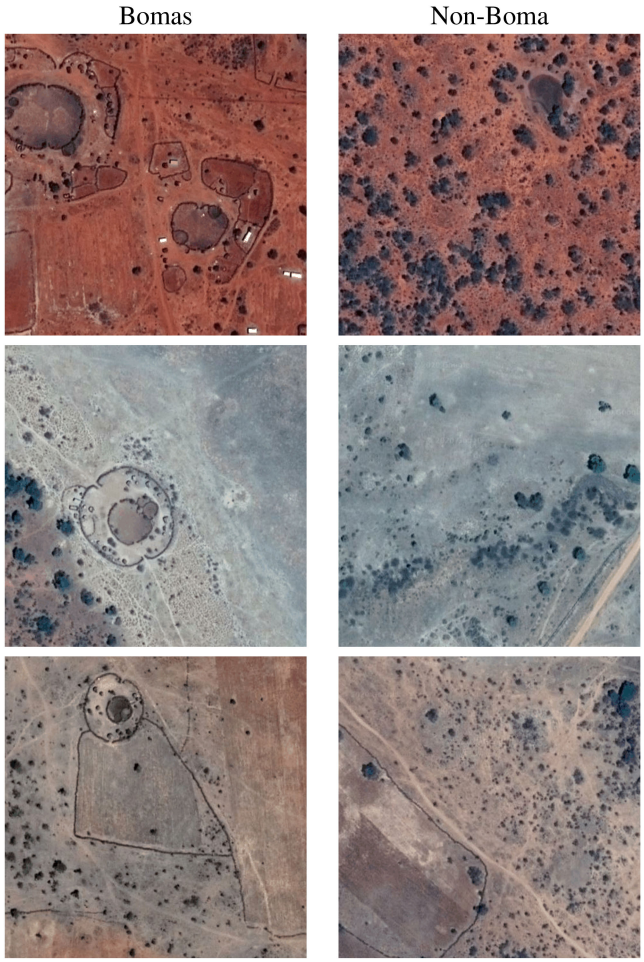


Fig. 3. Example image chips, 512×512 pixels. The left column contains positive (boma) training samples, discernable by the circular fencing lines and structures within. The right column are three negative (non-boma) training samples. Imagery copyright 2020, CNES/Airbus.

III. Automated Mapping With DNN

Automated mapping with DNN requires two key elements. First, a well-trained DNN that is robust to the target and environment, and also selective, i.e., good true positive rate (TPR) and low false negative rate (FNR). Second, it is necessary to be able to scale the application of the DNN over large areas, potentially in parallel, and then aggregate and refine the detections into a final ranked list for human analytical consumption. That is, the system should triage the area and rank likely boma higher and low-confidence boma lower. In [2], different DNN approaches were evaluated: the residual network, inception network, and neural architecture search (NAS). The ResNet50 [10] is a 50-layer deep vanilla residual network, while the Xception [8] is an extreme version of the inception module. The Inception-ResNet-v2 [9] combines the idea of residual network and inception module. The ProxylessNAS [7] is a NAS algorithm that searches without any proxy task and was proven to have better performance than other proxy-based NAS algorithms. These four architectures are representative, but not exhaustive, of the set respective DNN approaches. The top-performing DNN, ProxylessNAS (see Section IV-A) was then used for scanning the full study area.

A. Broad-Area Scanning

In the evaluation of automated boma mapping, the AOI was scanned with a 512×512 pixels window with a 512-pixel step size to get a sense of the performance of the models. However, during the analysis of the automated boma mapping results, a group of image chips with partial boma structure presented were falsely marked as non-bomas. To fix the false-negative issue and to geo-locate the bomas more accurately, the AOI was rescanned using an 8192×8192 pixels window with a stride of 32 pixels, 19.2 m. The overlapped tiles ensured that there was sufficient AOI coverage (no boma structures were partially captured on the edges). Each of the 8192×8192 pixels tiles was then sliced into 512×512 pixels chips. The chips were then passed to the pretrained ProxylessNAS models to generate geo-located classification response vectors. Fig. 4 shows the workflow from AOI tile to chip to DNN inference to classification vector. The response vectors were geo-located based on the latitude and longitude of the chip center.

B. Mode Seeking

The CVRF generated from the DNN can be sliced into two classification layers, one for bomas and one for non-bomas. For bomas, a response layer *R* can be sliced from the DNN classification responses to form a surface on the AOI whose topology is formed by the DNN output neurons. Regions with possible boma structures are expected to have peaks in *R*. The peaks in *R* are then ranked by their confidence scores to assist broad area search. To discover the peaks, an alpha-cut was applied using a confidence threshold. Since the final layer of DNN used the softmax function, the confidence scores range from 0.0 to 1.0, and the sum of network outputs per chip is 1.0. We used a threshold of 0.99 to dramatically reduce the data space to the most confident points of the DNN model.

In summary, we generated a density spatial field in which dense regions have a higher probability of having a boma structure. After that, the data points around the local peaks were clustered and the clusters were labeled and ranked to assist the human search process. As mentioned previously, the softmax classification in the last layer makes the sum of outputs equal to 1.0. Thus, when examining a single class from the responses individually, there will be irregularities and holes appearing in the surface. These irregularities and holes may separate a single large density into several small disparate densities. To reduce the effects of irregularities and holes on the response surface, an additive function-to-function morphology was applied to amplify the spatial density surface. For each data point, we define the spatial decay function $s(p) = exp^{(-d/D)}$ to describe the closeness between the data point and one of its neighbors, where d is the haversine distance between the two points and D is the radius of the neighborhood $N(p)$. Then, the point p's score in the amplified spatial density field is computed based on the intersected volume defined by the spatial neighbors of p and $N(p)$ [18], i.e.,

$$\delta(s(p), N(p)) = \sum (max(p, n) * s(p)). \tag{1}$$

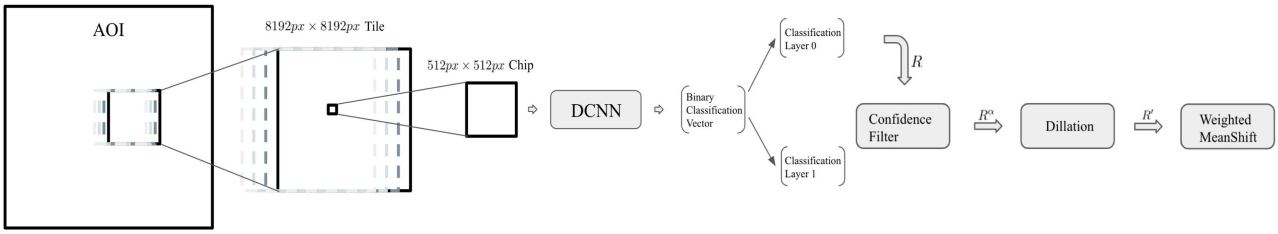


Fig. 4. Scanning workflow. The AOI is scanned as a set of overlapped tiles, then each tile is sliding-window processed into a set of 512×512 pixel chips. The chips are then pushed into the DNN models and produce the binary geo-located classification vectors.

Algorithm 1: Spatial Clustering of DNN Response Surface (R).

Input: Response Surface, R
Input: Movement Threshold, ϵ
Output: Mode-Clustered Response, R'

begin
  /* Alpha-cut, α = 0.99 */
  $R^\alpha \leftarrow$ confidenceFilter(R, α);
  /* Dilation of R^α by $s(p)$ */
  $R' \leftarrow \{\delta(s(p)), N(p)\}\, \forall p \in R^\alpha$;
  /* WeightedMeanShift */
  do
    $R'_{next} \leftarrow$ WeightedMeanShift $(R', s(p))$;
    $M \leftarrow$ haversine (R'_{next}, R');
    $R' \leftarrow R'_{next}$;
  while $M > \epsilon$;

Algorithm 2: Labeling Points as Clusters (R').

Input : Mode-Clustered Response, R'
Output: Ranked Clusters, C
/* $C_i \leq C_{i+1}$ */
begin
  $i \leftarrow 0$;
  while $R' \neq \emptyset$ **do**
    $p \leftarrow pop(R')$;
    /* Init. cluster C_i with chip p */
    $C_i \leftarrow p$;
    $N \leftarrow NearestNeighbor(p, R', D)$;
    foreach $n \in N(p)$ **do**
      $C_i.score = \{C_i, n\}$;
      $R'.remove(n)$;
    $C_i.score \leftarrow \sum_{c \in C_i} c.confidence$;
    $i++$;
  $\{C_i\} \leftarrow$ Sorted$(C_{\forall i})$ by score, descending

In this work, we applied the *weighted mean-shift* algorithm to find the clusters based on the amplified field of the DNN responses. The main advantage of the mode-seeking clustering algorithm is that its performance does not depend on a predefined number of clusters. For a field with amplified density, the clusters are the centers of mass of spatially connected densities. Since the number of such density centers was unknown, we applied the *weighted mean-shift* algorithm to find the modes. During each iteration, a spatial aperture of the nearest neighbors of a certain data point p were evaluated. Then the point p was moved to the center of mass of its spatial local neighbors.

Algorithm 1 describes the procedures to cluster the response surface R from DNN to mode-clustered response R'. We used a confidence threshold of 0.99 for alpha-cut and 100 m for the radius of spatial neighborhoods. First, a confidence filter was applied on the response surface R, only points with DNN's outputs confidence scores larger than 0.99 were selected for clustering. Then, the spatial decay function, $s(p)$, was used to weight the mean density of each local field.

The *WeightedMeanShift* function computes the weighted distance between the points in amplified response surface R' and its neighbors as $d(p, n) = haversine(p, n) * s(p)$. Then, the points are continuously evaluated and shifted toward the modes until the final movement is less than a certain threshold or the iterations are greater than the iteration limit. In the experiments, the movement threshold was set to 1×10^{-6} m and the maximum number of iterations was 500.

After shifting, data points were converged to the modes. Then, the modes were clustered and labeled using Algorithm 2. Each cluster was formed with a data point and its neighbors within neighborhood radius D. The scores of the clusters were computed as the volumes under their amplified spatial density. At the end of labeling, the clusters were ranked based on their scores for subsequent human analysis, and clusters with single-detection were considered as spatial outliers and filtered out of the result.

The *weighted mean-shift* algorithm becomes computationally expensive with large-scale datasets, however, the locality attributes of the CRVF allow a geometric decomposition and reassembly of a solution. To reduce the processing time, we designed a two-phase *weighted mean-shift*. During phase 1, the data points are divided into batches, *weighted mean-shift* is applied on each batch in parallel to generate a set of mini-clusters. In phase 2, these mini-clusters are considered as data points, and the scores of these mini-clusters are the confidence scores. Then *weighted mean-shift* is executed again to merge a certain range of mini-clusters into one large cluster. Combined with parallel computation, the modified *weighted average shift* significantly reduces the computation time for large-scale datasets.

IV. Experimental Results

To evaluate the effectiveness of DNN for detecting and mapping boma, we conducted a diverse set of experiments. First, we recap our closed-world model evaluations using traditional cross-validation techniques and traditional metrics. Then, we explore the performance and metrics in the nonclosed-world, real application scenario of automated mapping and scanning precision metrics. These extended experiments provide not just a broad understanding of the suitability of the DNN for this

TABLE I
NETWORK CROSS-VALIDATION PERFORMANCE ON MAASAI BOMA

Network	Prec.	Recall	F1	Error Reduction	Time (hrs)
ResNet50	95.374	95.381	95.322	-	8.572
Xception	95.893	95.339	95.450	2.744	9.856
Incept.ResNet-V2	96.834	96.780	96.780	31.180	12.696
ProxylessNAS	97.380	97.288	97.288	42.026	6.923

TABLE II
NORMALIZED CROSS-VALIDATION CONFUSION MATRICES FOR RESNET50, XCEPTION, INCEPTION-RESNET-V2, PROXYLESSNAS

Network	TPR	TNR	FNR	FPR
ResNet50	0.88	0.98	0.02	0.12
Xception	0.96	0.95	0.05	0.04
Incept.ResNet-V2	0.93	0.98	0.02	0.07
ProxylessNAS	0.95	0.98	0.02	0.05

particular task, but also a sense of how well the models transition to real-world applications such as broad area search.

A. Cross-Validation

As a baseline, we can look to the results from [2], summarized in Table I. The four evaluated networks were trained using 1×10^{-4} initial learning rate, the Adam optimizer, and batch size 8. To overcome the limited amount of training data (635 boma, 1726 non-boma), we leveraged transfer learning and augmentation as prescribed in [12]. We can see that ProxylessNAS achieved a recall of 97.288%, which is important for the mapping task. In the mapping scenario, the recall represents the likelihood of a false negative, i.e., a missed detection.

Table II shows that all the DNNs other than Xception achieved true negative rate (TNR) 0.98, on recognizing the non-boma images. In the broad area mapping use case, we want to minimize false negatives (FN) and will be more tolerant of false positives (FP), where the mode-seeking algorithms discussed in Section III help to filter the spurious FP.

B. Computational Costs

Not only did the ProxylessNAS architecture achieve the highest accuracy, but this model also consumed the shortest training time. Using the Nvidia Tesla P100 GPU, it took 8.572 h to train the ResNet50 model, 9.856 h to train the Xception model, and 12.696 h to train the Inception-ResNet-V2 model. In contrast, the training time for the ProxylessNAS model was only 6.923 h, which is $1.4\times$ faster than the Xception and $1.8\times$ faster than the Inception-ResNet-V2.

C. Automated Mapping Evaluation

To evaluate our trained ProxylessNAS in the broad area mapping task, we systematically downloaded 0.6 m nominal GSD images from the Google Maps Static API. The scanning evaluation area was bounded by 3° 58' 37" S – 4° 31' 18" S latitude by 36° 09' 05" E – 36° 43' 53" E longitude. Downloaded tiles were carved up into 512×512 px for batching into the DNN during inference.

TABLE III
SUMMARY OF SCANNING COUNTS FOR LEFT AND RIGHT

Model	Training Boma	Training Non-Boma	Testing AOI	Testing Tiles	Detection
L-trained	318	921	Right	20298	354
R-trained	320	805	Left	22089	425

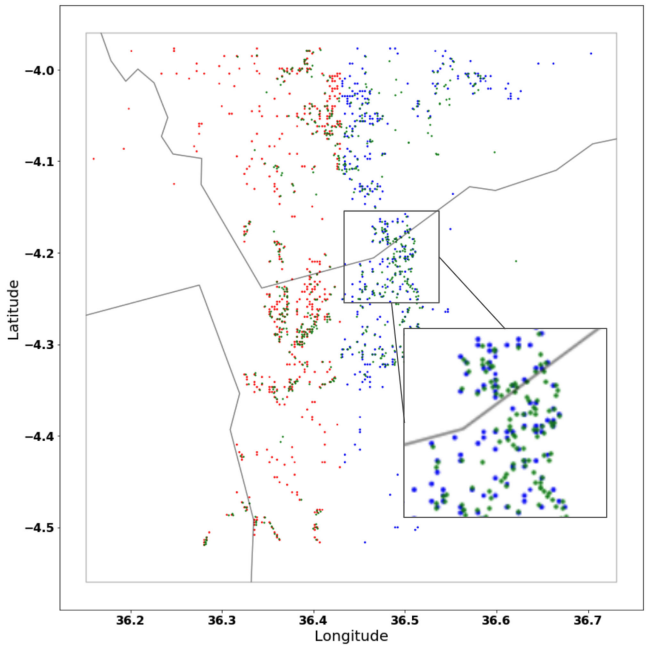


Fig. 5. Geospatial mapping of boma: Left trained ProxylessNAS shown as blue points on right of AOI, right trained network scan results shown as red points on left of AOI, ground truth bomas shown as green points.

We divided the ground truth into a *left* and *right* side at the median ground truth boma position of longitude (36° 25' 49.47656" E) to ensure blind testing evaluation. Two ProxylessNAS models were trained using data from the *left* and *right* sides, respectively. It took a total of about 4.97 h to train the two models using the Nvidia Tesla P100 GPU. These models were then used to scan the opposite side of the AOI. Scanning the left and right sides consumed 27 and 31 min, respectively.

Table III shows the counts for training chips for boma, non-boma, as well as total testing tiles, and the number of detections for each side. It can be seen that the total detections were 779, more than our ground truth, where we expect some FP, however, we see a high occurrence of human false negatives (HFNs). As noted in Algorithm 1, the confidence threshold applied to detections was 0.99.

Fig. 5 shows the respective detections, left-trained in blue and right-trained in red. As can be seen, the mapping of the detections closely follows the spatial pattern shown in the spatial training data (see Fig. 2). In fact, of the top 50 detections that were not colocated with known-ground truth, all were boma except ranks 24 and 25, which were abandoned boma. This speaks to two issues related to closed world evaluation. First, human-labeled data at a massive scale in the ever-changing landscape and environment of earth will always be missing true positive labels. Second, metrics that are more true to the real-world applications are needed to evaluate DNN *in situ* of the real-world use cases.

D. Scanning Results

The Google Maps Satellite images covering a 76.620 km^2 AOI across 4° 18.176' S - 4° 24.0813' S latitude and 36° 18.8062' E - 36° 22.5862' E degree longitude were acquired and scanned. A total of 127 008 768 chips were inferred using the ProxylessNAS models. After confidence filtering using a threshold of 0.99, the remaining 2485432 detections were used for dilation and spatial clustering. The processing time for each 8192×8192 pixels tile was about 5 sec using an NVIDIA GeForce GTX 1080 GPU. It took approximately 689 h to process the entire AOI, however, this can be accelerated with parallel computing techniques at the host-level.

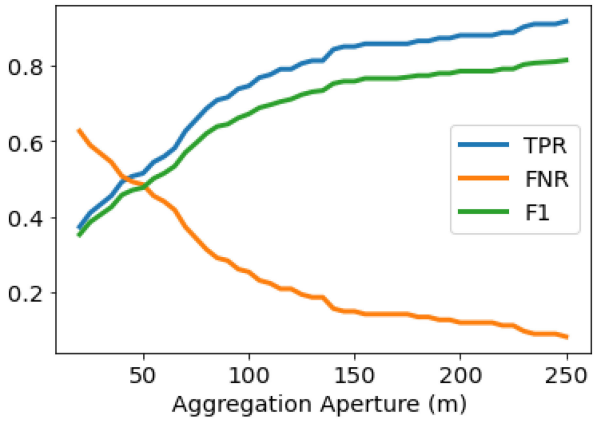
1) Scanning False-Positive Data Augmentation: An initial investigation was applied on the scanning and clustering results to find human false negatives. It used a batch size of 10, neighborhood radius D = 100 m, and searched 100 nearest neighbors in building the neighborhood during two phases of *weighted mean-shift*. During analysis of the scanning results, a few falsely detected bomas were found around the towns. Constructions in the towns have dense rectangular fencing structures that are easily confused with the boma structure. Thus, 27 additional non-boma examples were collected around three villages and added to the training dataset. A five-fold cross-validation experiment and automatic mapping evaluation were reapplied on the new training dataset. The F1-score achieved by the ProxylessNAS model was reduced from 97.288% to 96.337%, however, the mapping evaluation results showed those false detections around villages disappeared.

Fig. 6(a) shows the TPR, FNR, and F1 for scans using different aggregation apertures. At an aggregation aperture of 250 m, the model generated 194 clusters and achieved a TPR of 91.791% and an F1 of 81.485%. Ground truth derived FP generated using an aggregated aperture of 250 m were investigated for possible HFNs. Of the 65 candidates, 50 HFNs were identified. Fig. 7 shows the top ten clusters that went undetected during the human scanning. Most of the HFNs are small boma structures that are easily overlooked during human visual scanning. However, these examples received relatively high scores during the automated DNN and aggregation scan.

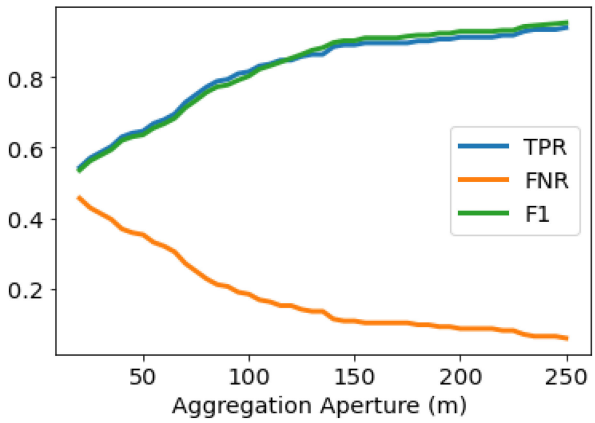
After adding the confirmed HFNs to the list of ground truth, the TPR increased to 94.022% and the F1 score increased to 95.395% using the aggregation aperture of 250 m. Fig. 6(b) shows the changes of updated TPR, FNR, and F1 scores with different aggregation aperture. As we see the F1 score lift above the TPR, we can infer that the recall or selectivity was also improved.

Borrowing from the information retrieval domain for scoring content-based query results [19], which is the analogous task of broad area search in imagery, we can use an order-dynamic retrieval precision as the *scanning precision*. Given a list L of ranked object candidates generated by Algorithm 2, then scanning precision for L is defined as

$$P_{\text{scan}}(L) = \sum_{i=1}^{m} \frac{i}{r(L,i)} \tag{2}$$



(a)



(b)

Fig. 6. Trends of true positive rate (TPR), false negative rate (FNR), and F1 relative to the aggregation aperture for ground truth scanning for both (a) before adding HFN as new TP to training, and (b) after adding HFN as new TP.

TABLE IV
DISTANCES BETWEEN DETECTED BOMAS AND THE GROUND TRUTH

	Mean	Min.	25%	50%	75%	Max.
meters	52.092	0.000	0.000	18.770	86.367	250.000

where $m = \min(|L|, n_r)$ and $r(L,i)$ is the rank of relevant (true) detection i in the set L. n_r is the number of true detections discovered *a posteriori* in L. Of note, the *a posteriori* is captured by m in (2), which can be set based on ground truth or an arbitrarily large value greater than the known ground truth. Fig. 8 shows the change of scanning precision and traditional precision with the results examined size. The difference between the scanning precision and traditional precision is more significant before the confirmed HFNs were added into the ground truth.

The effect of neighborhood radius during the second phase of *weighted mean-shift* was also investigated. Fig. 9 shows the change of TPR and F1 scores with different neighborhood radius and aggregation aperture. With a fixed aggregation aperture, the clusters produced using a smaller neighborhood radius lead to a higher TPR, while the F1 scores are almost identical.

Table IV summarizes the central tendency and dispersion of the distances between the positions of detected bomas and the centers of true bomas. The average distance between the

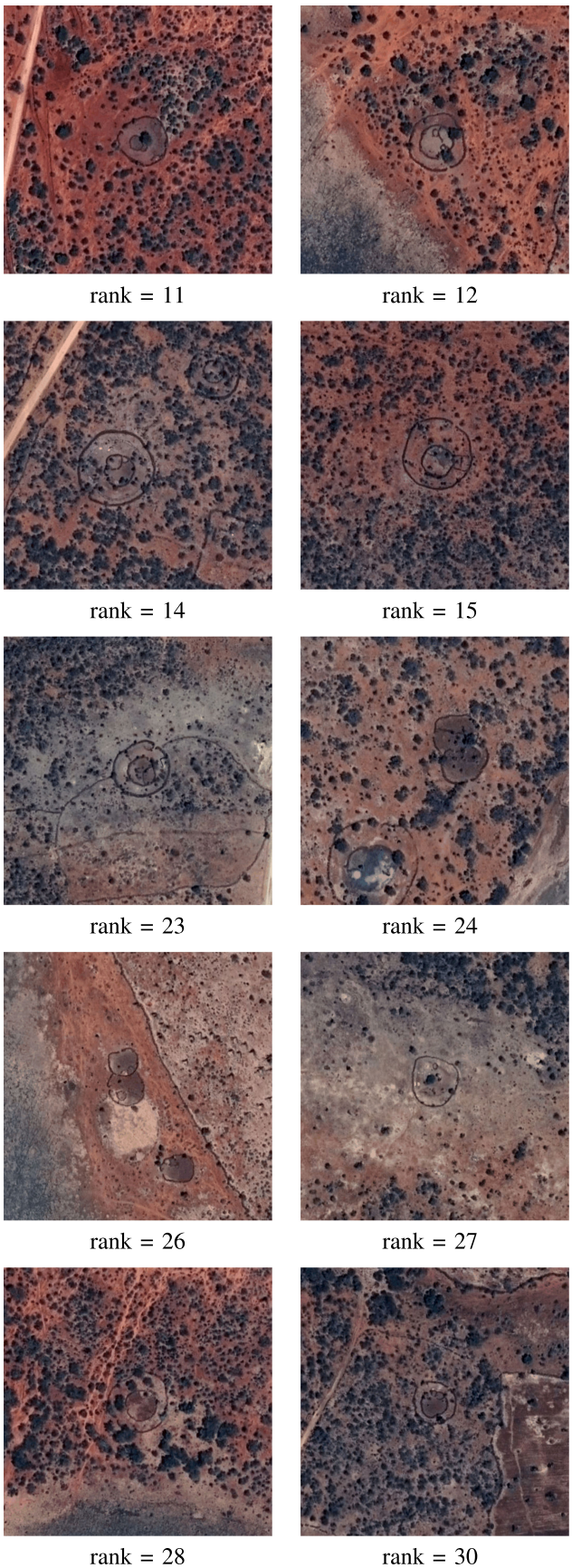


Fig. 7. Top ten human false negatives (HFN) during the sub-AOI scan. These clusters were ranked relatively high among the 194 clusters.

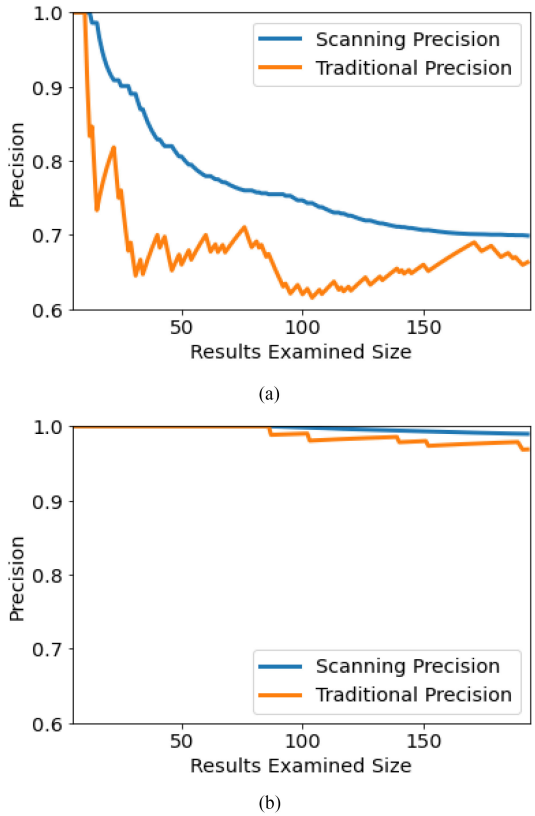


Fig. 8. Comparison of traditional precision and scanning precision (a) before adding HFN as new TP to training, and (b) after adding HFN as new TP.

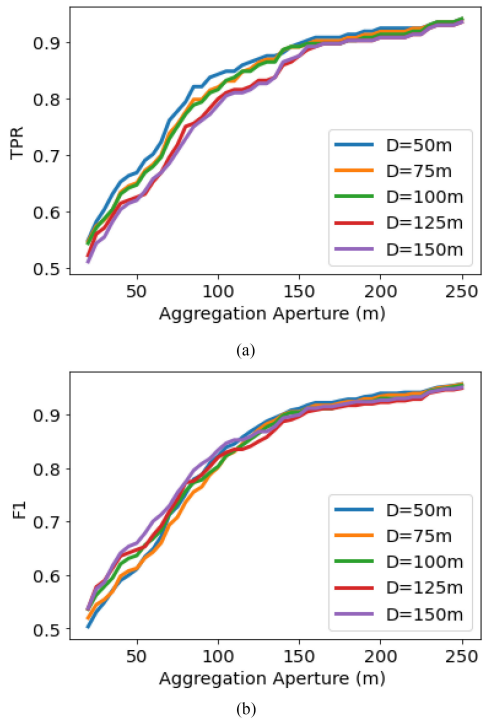


Fig. 9. Performance trends as the aggregation neighborhood aperture is adjusted during scanning results in (a) True Positive Rate (TPR) and (b) F1 score.

predicted positions of bomas and the centers of bomas is 52.1 m. The distances between 75% of the detected boma centers and true boma centers are less than 87 m. This tells us that using the DNN to perform broad area search and triage rank detections ensure that more than half the detections are within 19 m of a true boma.

V. Discussion

In this article, we have extended the investigations in [2] to the next logical step, that is evaluating the DNN in a real-world application scenario. We leveraged limited ground truth curated by human visual scanning in Google Earth Pro, along with RSI image chip downloads to train DNN and scan over 3900 km^2 in Tanzania. The DNN produced analytical layers have been given to the Humanity for Children CHAMP project to help bring humanitarian health care to the Maasailand population within Tanzania. In [2], we evaluated four representative DNN architectures for the boma-mapping task through five-fold cross-validation experiments. The ProxylessNAS model achieved the highest F1 score and run $1.4\times$ faster than the Xception and $1.8\times$ faster than the Inception-ResNet-V2. Due to its leading performance, the ProxylessNAS model was selected for automatic boma mapping.

A brief scan of the entire AOI was performed with a window of 512×512 pixels and a step size of 512 pixels. The detected boma distribution was highly correlated with the distribution of collected boma samples. Additionally, numerous *human false negative* (HFN) bomas from the human review process were detected with the assistance of the models. The brief scan is effective for roughly predicting the distribution of boma. However, in this process, some bomas may be missed due to the small scan window and large step size cutting off part of the boma structure at the edges. In addition, the positions of the bomas are imprecise. The distances between the roughly detected bomas and the real bomas may be around 300 m. Therefore, the results of the brief scan were not effective in helping CHAMP staff locate the bomas.

To geo-locate the bomas, we scanned a 76.620 km^2 sub-AOI using an 8192×8192 pixels window and a stride of 32 pixels. There were 50 HFNs (new discoveries) found among the 194 machine-detected bomas. Using an aggregation aperture of 250 m, the model achieved 94.022% TPR and 95.395% F1 score. The median distance between the detected positions and boma centers is 18.8 m. And for 75% of the detected bomas, the distances between their centers and the centers of true bomas are less than 87 m. With the scanning results, the time for ground survey can be greatly reduced and the CHAMP group is able to spread resources and clinics to more areas.

VI. Conclusion

In this article, we presented our work combining machine learning algorithms with remote sensing data to aid nonprofit organizations to map the Maasai boma homesteads. A complete procedure to map unmarked targets by scanning the earth with a trained deep neural network was proposed and evaluated. We presented novel adaptations for broad area scanning aggregations along with an in-depth analysis of parameterization of the aggregation. We evaluated the scanning results against the ground truth collected by humans and showed that the model was able to find the boma structures missed by human annotators. In addition, we reported quantitative metrics of scanning accuracy that allow interpretation of composite systems, deep neural models, and scan aggregation for human practitioners using this technique to aid in visual analysis tasks. This work addressed the challenge of utilizing deep neural networks or any applied machine learning models in practical applications of remote sensing, where absolute ground truth is difficult to achieve on the ever-changing Earth's surface.

There are many areas in the world, such as Tanzania, that lack resources and are not integrated into the existing systems. This approach can be applied to similar areas and assist organizations in locating remote villages, rare structures, or any variety of diverse anthropogenic features.

Scanning the area with a large window size and small stride is necessary for precisely locating the positions of bomas. This results in a computationally costly process using high-resolution imagery, even with the efficient ProxylessNAS model. One direction for future work is to optimize the scanning workflow using parallel computing and other techniques. Another direction is to train a robust model for lower resolution imagery, so that images covering the AOI can be replaced with lower resolution imagery and scanning time is reduced.

Acknowledgment

The authors would like to thank R. Hansen, J. Antel, and other members of the Community Health Assessment Mapping Project (CHAMP) by *Humanity for Children* and the Center for Health Analytics for National and Global Equity.

References

[1] T. Pheage, "Dying from lack of medicines," 2017. [Online]. Available: https://www.un.org/africarenewal/magazine/december-2016-march-2017/dying-lack-medicines

[2] K. Cheng, I. M. Popescu, L. Sheets, and G. J. Scott, "Automatic Maasailand Boma mapping with deep neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 2839–2842.

[3] J. A. J. Berni, P. J. Zarco-Tejada, L. Suárez, and E. Fereres, "Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 722–738, Mar. 2009.

[4] W. Tyler *et al.*, "Exploring the effects of class-specific augmentation and class coalescence on deep neural network performance using a novel road feature dataset," in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop*, 2018, pp. 1–7.

[5] V. L. Mulder, S. D. Michael, E. B. Schaepman, and T. R. Mayr, "The use of remote sensing in soil and terrain mapping-A review," *Geoderma*, vol. 162, no. 1/2, pp. 1–19, 2011.

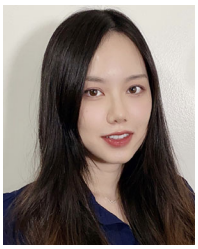
[6] L. Zhong, L. Hu, L. Yu, P. Gong, and G. S. Biging, "Automated mapping of soybean and corn using phenology," *ISPRS J. Photogramm. Remote Sens.*, vol. 119, pp. 151–164, 2016.

[7] H. Cai, L. Zhu, and S. Han, "ProxylessNAS: Direct neural architecture search on target task and hardware," *CoRR*, vol. abs/1812.00332, 2018.

[8] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1800–1807.

[9] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," *CoRR*, vol. abs/1602.07261, 2016.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[11] G. Cheng, C. Ma, P. Zhou, X. Yao, and J. Han, "Scene classification of high resolution remote sensing images using convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 767–770.
[12] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land-cover classification of high-resolution imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, Apr. 2017.
[13] N. Kussul, A. Shelestov, M. Lavreniuk, I. Butko, and S. Skakun, "Deep learning approach for large scale land cover mapping based on remote sensing data fusion," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 198–201.
[14] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Urban land cover classification with missing data modalities using deep convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 6, pp. 1758–1768, Jun. 2018.
[15] J. A. Hurt, D. Huangal, C. H. Davis, and G. J. Scott, "A comparison of deep learning vehicle group detection in satellite imagery," in *Proc. IEEE Int. Conf. Big Data*, 2019, pp. 5419–5427.
[16] S. Jiang *et al.*, "An optimized deep neural network detecting small and narrow rectangular objects in Google Earth images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1068–1081, Mar. 2020.
[17] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, "Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," *CoRR*, vol. abs/1811.12231, 2018.
[18] G. J. Scott, J. A. Hurt, R. A. Marcum, D. T. Anderson, and C. H. Davis, "Aggregating deep convolutional neural network scans of broad-area high-resolution remote sensing imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 665–668.
[19] G. Scott and C.-R. Shyu, "Knowledge-driven multidimensional indexing structure for biomedical media database retrieval," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 3, pp. 320–331, May 2007.



Keli Cheng (Student Member, IEEE) is currently working toward the Ph.D. degree in computer science with the University of Missouri, Columbia, MO, USA.

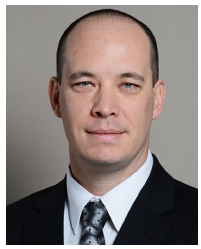
She is currently a Research Assistant with the High-Performance Data-Intensive Computing Systems Laboratory, University of Missouri-Columbia. She is exploring a range of image processing and information mining algorithms, including deep learning systems, to process remote sensing data into large-scale geospatial data cubes. Her research interests include deep learning, computer vision, large-scale data systems, and combining visual features with advanced spatio-temporal algorithms to perform predictive analysis of man-made and natural events in time-series images.

Ilinca Popescu is currently working toward the B.S. degree in earth systems program for the land systems track with Stanford University, Stanford, CA, USA.

Her research interests include hydrogeology, sustainable practices for future land use, and water equity.

Lincoln Sheets received the M.D. degree and the Ph.D. degree from the University of Missouri in 2011 and 2017, respectively.

He is a Co-Founder and one of the directors of the Center for Health Analytics for National and Global Equity (CHANGE), Columbia, MO, USA. He is a Physician and Medical Researcher who works as an Adjunct Professor of Health Informatics with the University of Missouri School of Medicine, Columbia, MO, USA. He has published, presented, and taught in the areas of geospatial information systems, global health disparities, and social determinants of noncommunicable disease. He gained 20 years of experience in software engineering before earning an MD and Ph.D. in Public Health Informatics, and he has volunteered in public advocacy and legislative lobbying locally and nationally. As a medical student, he organized medical missions in Africa and Latin America, and helped found a student-run free clinic in Columbia, MO, USA. His free-time interests include foreign languages, triathlon, mountaineering, cinema, and spending time with family at home in Columbia, MO, USA and his hometown of Springfield, MO, USA. His research interests include combining evidence-based medicine with information technology and business strategy, to improve health outcomes in underserved rural and minority communities of the United States and around the world.



Grant J. Scott (Senior Member, IEEE) received the B.S. and M.S. degrees in computer science and the Ph.D. degree in computer engineering and computer science from the University of Missouri, Columbia, MO, USA, in 2001, 2003, and 2008, respectively.

He is a founding Director of the Data Science and Analytics master's degree program with the University of Missouri, where he is an Assistant Professor with the Electrical Engineering and Computer Science Department. He is exploring novel integrations of computational hardware and software to facilitate high performance advances in large-scale data science, computer vision, and pattern recognition. He has leveraged his experience in the development of innovative remote sensing (satellite and airborne) change detection technologies, resulting in five U.S. patents. He has participated in a variety of professional networking and academic events, as well as worked with a variety of groups to bring data science training to their people (MU Public Policy, USDA, IEEE international conferences). His current research interests include real-time processing of large-scale sensor networks, parallel/distributed systems, and Internet of Things (IoT) data; deep learning technologies applied to geospatial datasets for land cover classification and object recognition; extensions of enterprise RDBMS with HPC coprocessors; crowd-source information mining and multimodal analytics; high performance and scalable content-based retrieval (geospatial data, imagery, biomedical); imagery and geospatial data analysis, feature extraction, object-based analysis, and exploitation; pattern recognition databases and knowledge-driven high-dimensional indexing; and image geolocation.