



OPEN

DATA DESCRIPTOR

ORBITaL-Net: A labeled training library for large-scale building feature extraction

Benjamin Swan ^{1,3} , Joe Pyle^{1,3}, Darrell Roddy^{1,3}, H. Lexie Yang ^{1,3}, Amy Rose^{2,3} & Melanie Laverdiere ³

Over the course of several years, nearly 1.5 million building outlines have been created from approximately 128,000 training tiles covering roughly 7,000 km² of very high-resolution multispectral overhead imagery, primarily dated between 2010 and 2020. This dataset, dubbed the Oak Ridge Building Image and TrAining Label Net (ORBITaL-Net), is designed for machine learning applications and is global in scope, with samples drawn from 72 countries across North America, South America, Africa, Europe, and Asia. ORBITaL-Net captures a great diversity in geographic setting, structural characteristics, land use (urban and rural), terrain, and imagery conditions. While the labeled building outlines are themselves valuable, the dataset's true strength lies in the pairing of these labels with corresponding reference imagery, which is being released for open source use. Similar to SpaceNet and Replicable AI For Microplanning (ramp), this building outline dataset will allow the larger computer vision community from academia, government, and industry the opportunity to develop robust, scalable, and generalizable geospatial machine learning techniques. Unlike SpaceNet and ramp, which offer high resolution labels and imagery primarily for large urban cities, ORBITaL-Net is not focused on training samples from heavily populated areas but instead aims to capture the innate variability of conditions present in both the physical environment and imagery collections.

Background & Summary

Background. In the last decade, the convergence of big data, automation, and artificial intelligence (AI) has made the possibility of building-level mapping at global scale a reality. In step with these technological advancements, the value of large-area building mapping continues to grow across sectors. Government agencies and NGOs rely on accurate locations and descriptive information about buildings and related occupancy for emergency preparedness, response, and recovery actions associated with natural and anthropogenic hazards¹⁻³. Within urban planning, real estate, and insurance industries, detailed spatial and attribute information informs mobility⁴, site selection⁵, and risk applications⁶. Climate and energy are intrinsically linked, and the morphological and functional aspects of the built environment play an important role in understanding the role of energy efficiency on future climate impacts and mitigation strategies^{7,8}. Across all sectors, practitioners are moving toward creating and enriching digital twins with increased realism, to improve Building Information Models⁹, to simulate urban dynamics^{10,11}, and to test what-if scenarios in a safe digital space¹².

A specific catalyst for the work described here was the requirement to map any building(s) in the world to be used as foundational input to geospatial models of human settlement and associated population counts, particularly the Oak Ridge National Laboratory (ORNL) LandScan family of gridded population data. Initially, a support vector machine (SVM) method was employed to extract settlement features with an output resolution of 8 m from high resolution imagery^{13,14}. However, SVMs do not generalize well to large, diverse target areas^{15,16}. As such, training data and models were developed per image scene, which becomes impractical for global-scale mapping applications. As an example, hundreds of models were developed in order to map all the settlement areas for the country of Nigeria. Further, in many cases the output resolution of 8 m was too coarse to adequately represent between-building spacing and variability of this factor between settlements. Although this approach

¹Oak Ridge National Laboratory, Geospatial Science and Human Security, Oak Ridge, 37830, Tennessee, USA.

²Overture Maps Foundation, Oak Ridge, USA. ³These authors contributed equally: Benjamin Swan, Joe Pyle, Darrell Roddy, H. Lexie Yang, Amy Rose, Melanie Laverdiere. ✉e-mail: swanbt@ornl.gov

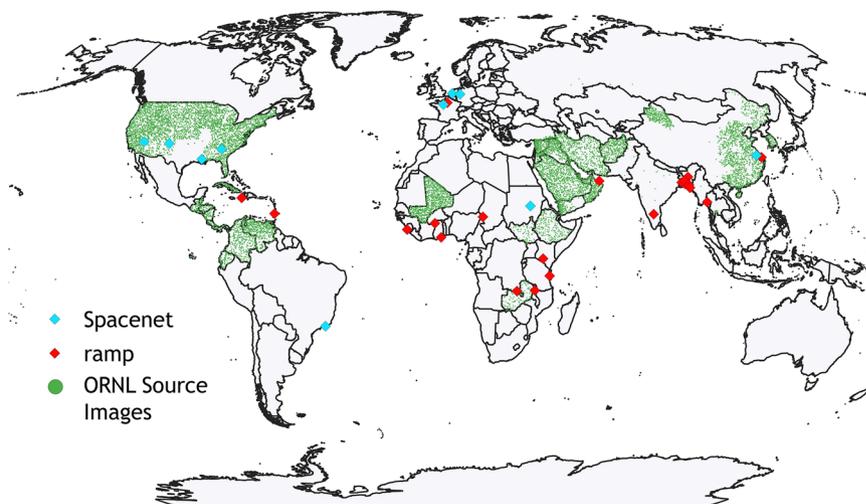


Fig. 1 Comparative distribution of source image locations for the ORNL, SpaceNet, and ramp open building training datasets.

could still produce a high quality “footprint” of human settlements, this variability had a non-negligible impact when developing estimates of population within those settlements.

With the maturation of deep learning for remote sensing applications^{17–19}, we moved to a Convolutional Neural Network (CNN) method for building extraction²⁰. Unfortunately, a significant amount of labeled training data is required to learn an accurate model. SpaceNet²¹ and Replicable AI For Microplanning (ramp)²² offer high resolution labeled training data and accompanying reference imagery for 10 and 22 population centers, respectively. However, these datasets are partial to heavily populated areas. To work toward a model that could be generalizable to the variety of geographies across the globe, a robust training library needed to capture that variability in terms of structural, terrain, and imagery characteristics. To date, there are no open-source datasets that can provide the quantity, quality, and variability available in the labeled training data library that is presented here. ORBITaL-Net²³ consists of 128,000 500×500 pixel chips and 1,490,000 hand-labelled buildings drawn from 9,800 distinct high resolution satellite images. Figure 1 shows the spatial distribution of this dataset along with SpaceNet and ramp.

Toward a generalizable model. For the last decade, CNNs have been at the forefront of computer vision and now enable the creation of highly capable remote sensing feature extraction models that can operate over complex datasets of a size impossible using earlier techniques^{24–26}. However, the generalization ability of a CNN model depends largely on the amount and quality of the training data used in supervised learning settings²⁷. More complex CNN models tend to result in better overall performance, however, the number of parameters also increases as the complexity of a deep feature learning models grows. This demands more labelled training samples in the model optimization and learning process. The remote sensing community is aware of the need for an open, comprehensive labelled training dataset to leverage the advances made in artificial intelligence (AI) and enable large-scale mapping with high resolution earth observation data. As a result, a number of community collaborations and open data initiatives have led to the development and availability of several impactful labelled training datasets, such as Functional Map of the World²⁸, SpaceNet Challenges²⁹, BigEarthNet³⁰, xView Challenges^{31,32}.

Training CNN models with such a massive amount of data is also computationally expensive, which largely limits the impact of those labelled data. The transfer learning strategy has been adapted, to take advantages of the pre-trained models without duplicating the efforts and time to build a model from scratch. The labelled data we created in this work provides a base model for building mapping for anywhere in the planet.

Supervised building extraction models could be over-fitting to the training set. For example, we observe that a given trained model can perform very well on a set of training labelled samples, however the model fails to generalize well (i.e. providing good or comparable accuracy on new/unseen labelled samples) as the models only memorize the training samples rather than learn the underlying feature patterns. With the goal of creating a generalizable model for extracting buildings across very large geographic areas from high-resolution satellite images, we recognized the need to build a labelled data set to cover variability in building appearances and image characteristics. These differences often present challenges to a pre-trained model as these models do not often “see” the rare representations of buildings and domain shifting of the pixel values under the impact of various imaging conditions. Diversifying training data has been effective in improving model generalizability. This strategy can expose those edge cases to the models during training and optimization and reduce the model bias. This strategy can also help fit models with an enormous number of parameters, as recent studies show that larger models with higher capacity need abundant data to reach their full potential³³. The performance gains from increasing model size diminish without sufficient training examples.

Diversity of training samples. In order to ensure enough positive sample variability existed in the labelled training set, a variety of buildings with varying uses (e.g. residential, commercial, industrial), construction

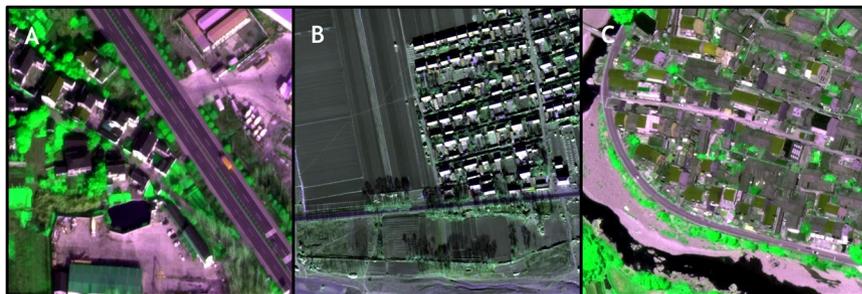


Fig. 2 Examples of challenging variation in image acquisition conditions included in our dataset (A) High off-nadir view angle, mid sun elevation (B) Low off-nadir angle, low sun elevation (C) Mid off-nadir angle, high sun elevation.

materials (e.g. mud bricks, steel, wood, straw), and shapes (e.g. square/rectangular, circular, irregular) were identified and introduced to the model. With sufficient quantity, this variability increased generalizability across large, diverse areas. As an example, a single model was used to automatically extract buildings from the countries within the Arabian Peninsula—an area of approximately 3.1 million sq km.

Core to our direction of producing a globally generalizable model is the development of a training dataset that includes variability in both image acquisition contexts and vernacular architecture and cultural settlement patterns. Key variables that can affect the appearance of include the viewing angle of the image sensor, which affects the amount of building facade visible, and the sun elevation, which affects scene contrast and presence and degree of shadows. Some challenging image acquisition conditions are illustrated in Fig. 2. The considerable variation in architecture, settlement configuration, and land cover contexts within a single country is illustrated in Fig. 3, which highlights three of the numerous regions of Ethiopia. On the left hand side of each image pair is a training sample and on the right the output from our current global model for a nearby area.

In addition to showing the wide spatial distribution of ORBITaL-Net in Fig. 1, we have compiled basic geometric statistics on the building outlines as well as their environmental contexts. Shown in Table 1 are the central tendencies and standard deviation in four variables: area, perimeter, complexity (the ratio of area to perimeter), and isoperimetric quotient³⁴, a measure of how close to a circle a shape which is calculated as:

$$ISP = \frac{(4 * \pi * area)}{perimeter^2} \quad (1)$$

Figure 4 then shows the percentage of buildings that fall within a given biogeophysical zone, as defined by the World Wildlife Federation and Fig. 5 the second level Koppen climate zone.

Impacts of the developed training dataset. As labelled data continued to be added to the library, less variation between the pre-trained model and the resulting fine-tuned model was noticed, leading to fewer and fewer samples generally being needed to adapt the model to a new project area of a given size. For instance, between the Arabian Peninsula, a project area we used early in the compilation of our training library and a far later project area of Central and northwest South America, the labelling effort in terms of work hours decreased by almost half. Since we have completed the full, global training dataset, and used it to train a Global Base Model we have eschewed the need for obtaining additional labelled data in most circumstances. Examples of areas where we have successfully used our Global Model with no fine-tuning include Taiwan, Sudan, Ukraine, Poland, and Belarus.

To illustrate the improvement in generalizability of our models over time, as a greater number and variety of training samples added, we performed a retrospective validation of previously trained and fine-tuned models used for building layer production against a set of 9,500, globally distributed validation samples. For this analysis, we used three standard metrics: (1) precision, which measures how reliable predictions are, (2) recall, counterpart to precision that measures how complete predictions are, and (3) F1-score, the harmonic-mean of precision and recall which can be used to balance tradeoffs between these two metrics. These are calculated as:

$$precision = \frac{tp}{tp + fp} \quad (2)$$

$$Recall = \frac{tp}{tp + fn} \quad (3)$$

$$F1 = \frac{2 * precision * Recall}{precision + Recall} \quad (4)$$

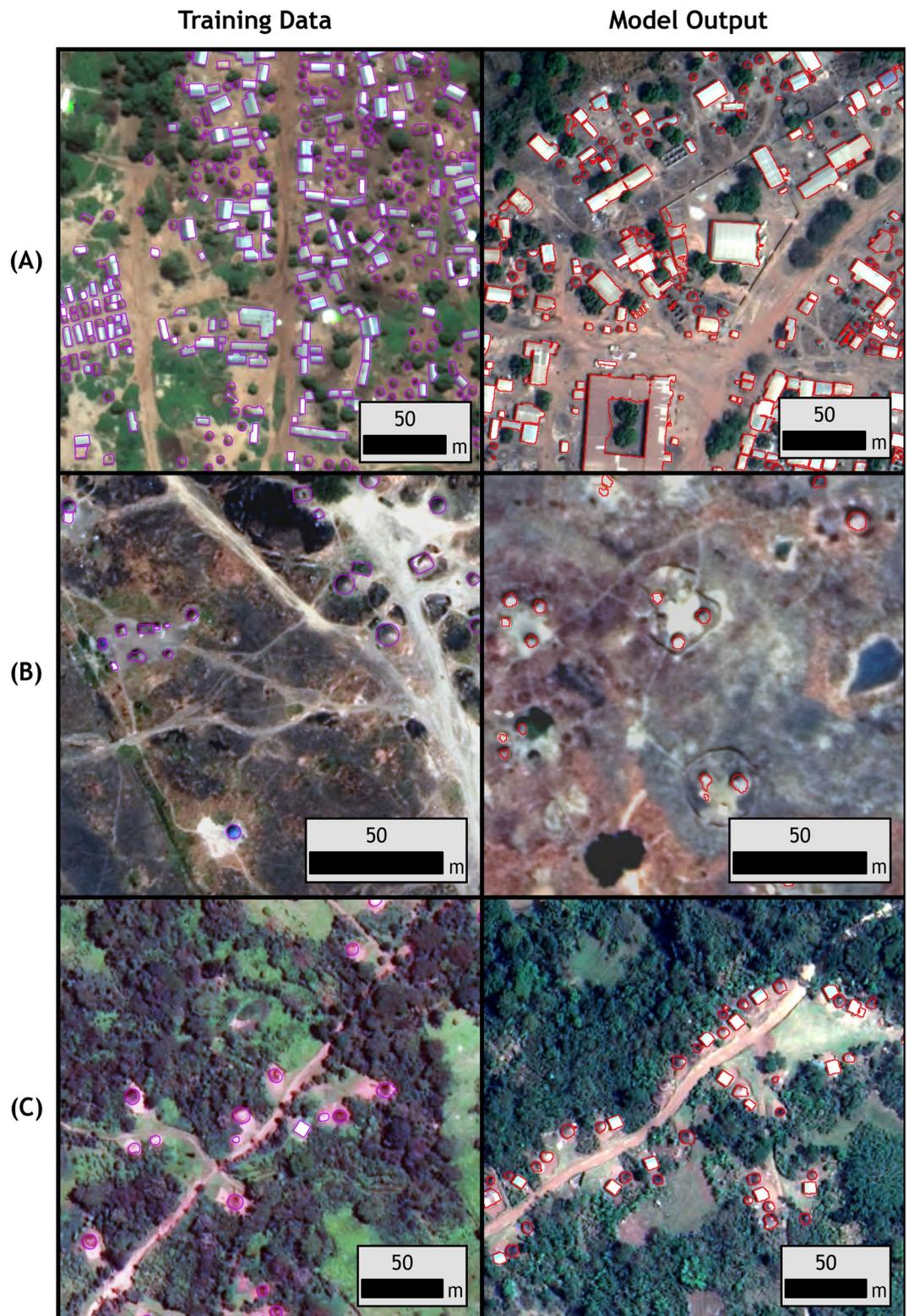


Fig. 3 Illustration of just a few examples strong regional variation in building style and context that can be found across an area like North Africa. For each pair, left shows a single sample from our database, right the output of our global model for a nearby area. Illustrated: (A) Sikasso Province, Mali (B) Jungoli Province, South Sudan (C) Bench Sheko Zone, Ethiopia. ©Maxar 2023.

The results are plotted in Fig. 6. This is presented in chronological order, from the initial project areas of North Korea and Yemen all the way to the current, Global Base Model trained on our full dataset. There is a clear trend over time (and thus expansion of the global training dataset in quantity and variety of samples) towards

Metric	Mean	Median	Std
Area m^2	140	97	146
Perimeter m	47	4	27
Complexity	0.98	0.44	0.44
IPQ	0.68	0.7	0.13

Table 1. Descriptive Statistics for the structure polygons in the dataset (excluding structures $>1,000 m^2$).

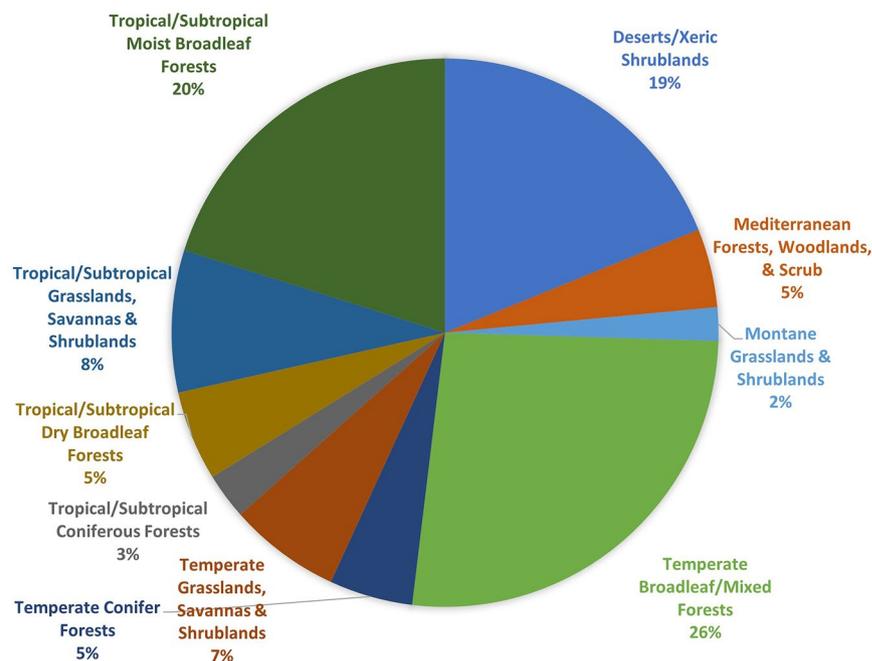


Fig. 4 Percentage of buildings in the dataset within each biogeophysical region as defined by WWF.

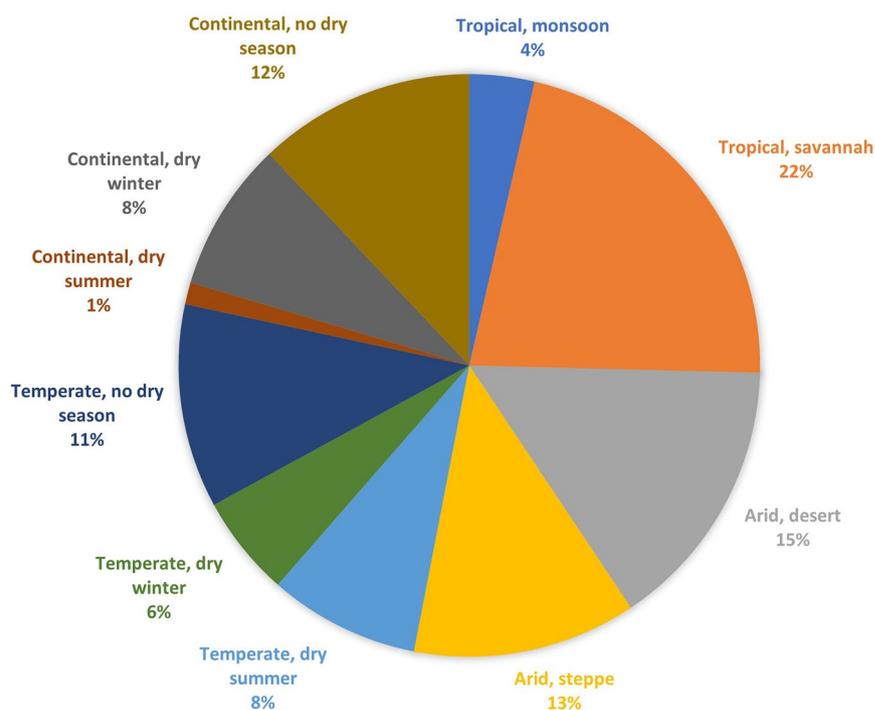


Fig. 5 Percentage of buildings in the dataset within each level 2 Koppen climate zone.

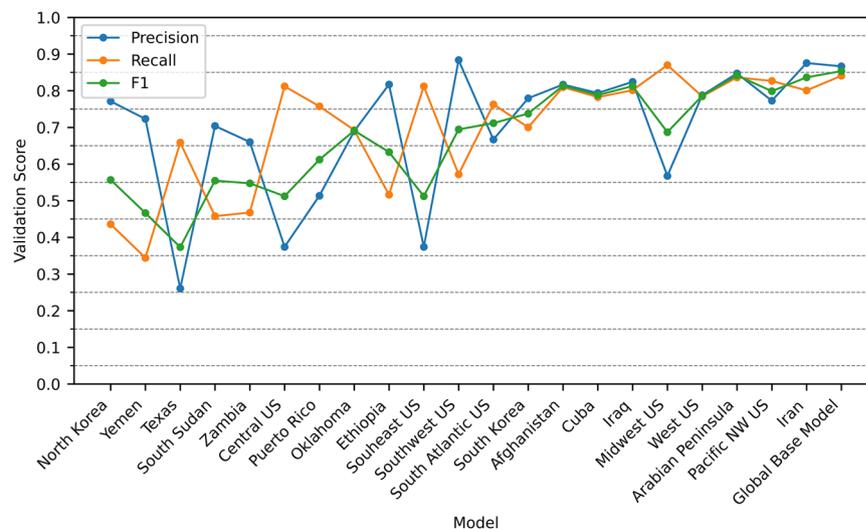


Fig. 6 Performance of ORNL building segmentation models, fine tuned for project areas, measured on a set of 9,500 withheld samples drawn proportionally from imagery across our global dataset. This presentation in order of recency (left to right) shows the steady improvement of the generalizability of our fine-tuned models as a more extensive and diverse training dataset is used.

models that are more generalizable according to both the increase in overall F1 score and the narrowing of the ratio between precision and recall scores, a general indicator that a model has learned a more “stable” prediction function.

Summary. We summarize ORBITaL-Net with a listing of what we consider it’s most salient strengths and weaknesses.

Strengths

- A broad geographic coverage including a variety of architectural styles, settlement patterns, and biogeophysical contexts across four continents.
- Incorporation of many medium and low population areas rather than a focus specifically on major urban areas.
- Inclusion of a large number of pure negative (no buildings present) samples, which we have found to be very beneficial for obtaining a good model precision across large, diverse geographic areas.
- Use of a large number of individual source satellite images that better represent the variety of image acquisition contexts seen in real-world applications.

Limitations

- The source imagery used is exclusively from Maxar very-high resolution sensors (primarily WorldView-2 and WorldView-3); therefore, a model trained using this dataset may have considerably worse performance when used with image data that are substantially different to these without the use of transfer learning.
- As we labelled all visible parts of the building as such, if a user wished to have a different definition, such as extracting rooftops only, the utility of this dataset may be lower.
- Due to the fact that these samples were designed to train a model to extract all visible parts of a building instead of the footprint, there may be discrepancies between the area of building features predicted by such a model and the true building footprint area, particularly when working with imagery at a high off-nadir angle and/or very tall buildings. This can have implications for such applications as assigning population by building area.

Methods

ORBITaL-Net was created over a span of five years (2017–2021) through a series of large-scale building extraction projects. Project areas were selected to be generally homogeneous in terms of climate and architecture/settlement patterns and ranged in geographic coverage from a cluster of provinces within a single country to multi-country regions. For each project area, we created a set of hand-labelled samples for fine tuning our pre-existing base building segmentation model, the first iteration of which was described in²⁰. These new samples were added to the full training pool, increasing the overall generalizability of our base model.

Overhead imagery was sourced from Maxar and consists primarily of 4-band (red, green, blue, and near infrared bands) images collected from the WorldView-2 and WorldView-3 satellites. Images from the

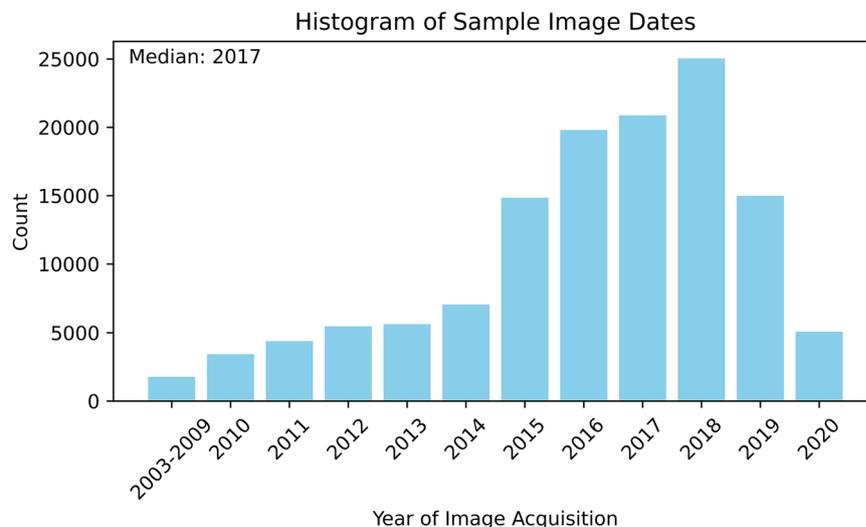


Fig. 7 Histogram of all individual samples by year of source image acquisition.

QuickBird-2 and GeoEye-1 satellites were used where suitable WorldView imagery was not available. The pan-sharpened resolution of the resulting image collections is 0.45 m with a minimum acceptable resolution of 0.60 m and maximum off-nadir view angle of 40 degrees. The median resolution for samples in the building dataset is 0.47 m with a standard deviation of 0.09 m. We received this imagery at Level-1B (sensor corrected, un-projected (raw) imagery product) and applied our in-house georegistration, orthorectification, and pan-sharpening, as described in³⁵. The bulk of the imagery dates between 2015 and 2020, with a median acquisition year being 2017, as shown in Fig. 7.

Once the source imagery was pre-processed, the next step was identifying a suitable set of samples for labeling. To address the challenge of quickly identifying a highly salient set of samples across a large set of imagery with multi-modal variability, we developed the ISOSCELES program. This is a fully automated program, described in more detail in³⁶ that performs a multi-level unsupervised clustering to select first the most representative satellite image scene from the full dataset and then the most representative chips (image subsets at the scale of the inference model the user chooses) within each scene. Features included for the selection of satellite images included sun elevation, image sensor viewing angle, and overall spectral statistics. In turn, image texture and spectral statistics are used to aid in the selection of representative image chips.

This set of chips was then divided into positive (buildings present) and negative (no buildings present) samples. In accordance with the general practice of using training data sets balanced between classes, we have found across numerous project areas that a mixture of 55–60% negative to 45–40% positive samples provides the best trade-off between false positives and false negatives.

Hand-labelled training data. Our initial proof of concept of the feature extraction at scale²⁰ utilized existing LiDAR building footprints to generate labelled samples. While that was a successful demonstration, we identified the model generalization issue when we investigated the feasibility of adapting the learned building extraction models to other countries. The main challenges stem from the differences in images at various aspects, including the inferior model performance when the image quality is poor, or variation in structure types due to cultural or socioeconomic differences. To address this issue, we decided to use a domain adaptation approach to fine-tune our best performing model to each new target project area. However, we quickly discovered that the lack of reliable labeled data became a significant barrier to this approach. Without labeled data that is both diverse and reliable, it's difficult to improve the model's ability to generalize to new environments.

A team of 30 analysts leveraged GIS software to digitize features of interest. Analysts were trained how to distinguish building structure features from non-building features in imagery with emphasis placed on consistent, uniform digitization of building features.

Each building was digitized as a discrete, non-overlapping polygon (Fig. 8), with all visible portions of the building labelled, including facades (Fig. 9). Imagery properties (brightness, contrast, stretch, band combinations, etc.) were adjusted for each image chip so all buildings in the image could be positively identified. Some examples are shown in Fig. 10.

Our definition of “building” was guided by the general use-case of producing building datasets that could be used as foundational spatial inputs for population modelling. Thus, the rule-of-thumb was if a person might plausibly occupy the structure, be it as a dwelling, workplace, school, store, etc., it would be considered a building. Table 2 gives a more detailed description what features were digitized as buildings, what features commonly found on roofs were also labelled as building, and which human-built features were not labelled as buildings. In cases where portions of buildings were obscured by shadow or vegetation, those pixels were labelled as a building if there was partial occlusion and non-building if they were fully occluded.

Variation in how buildings were digitized stemmed from multiple facets of imagery interpretation, including architectural differences (Fig. 11) and challenging image acquisition conditions (Fig. 12). Defining the outline



Fig. 8 Building features were digitized as non-overlapping polygons. (a) sample image. (b) sample image with labels. ©Maxar 2023.



Fig. 9 Building facades were included in labeling. (a) image. (b) image with labels. ©Maxar 2023.

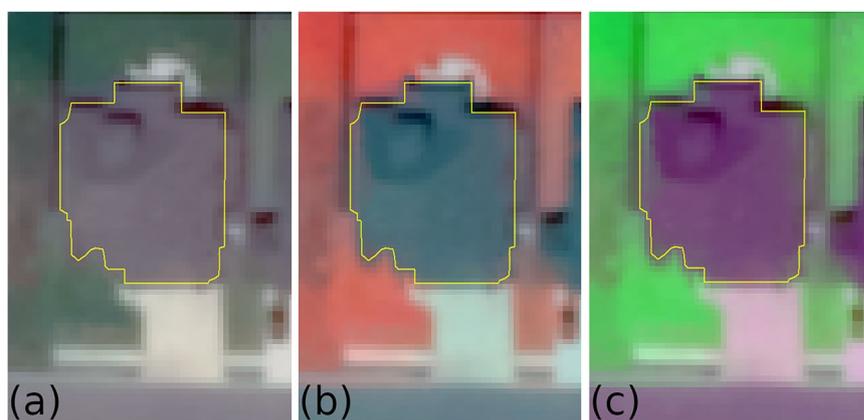


Fig. 10 Different band combinations were used to help differentiate buildings from vegetation. (a) shows the RGB combination of bands 3,2,1 which is true color, (b) shows the RGB combination of bands 4,2,1 where the vegetation appears red, and (c) shows the RGB combination of bands 3,4,1 where vegetation appears bright green. ©Maxar 2023.

of a building requires an experienced analyst to resolve the artifacts of view angles (e.g. shadows), identify salient morphology features (e.g. walled boundaries between adjacent buildings, non-permanent features such as

Category	Description
Structures Labelled Building	Structure with a complete roof, huts, carports, covered porches and walkways, silos, greenhouses, and building facade.
Roof Items Labelled Building	Solar panels, antennas, chimneys, smokestacks, HVAC, skylights.
Structures Not Labelled Building	Buildings without a roof, buildings under construction, damaged/destroyed buildings, vegetation occluding a building, courtyards, bridges, dams, parking garages (elevated parking lots), communication towers, transmission line towers, substations, stand alone solar arrays, solar farm inverters, windmills, canopies, tents, tanks (horizontal and vertical) for fuel/gas storage, cemetery features (mausoleum, grave markers, etc), elevated pipes, frameworks, gantries, ground-mounted electrical transformers, agricultural row covers, vehicles, trailers, and storage/shipping containers.

Table 2. Details of the rubrics used to determine whether an object in an images should be labelled as building or non-building.

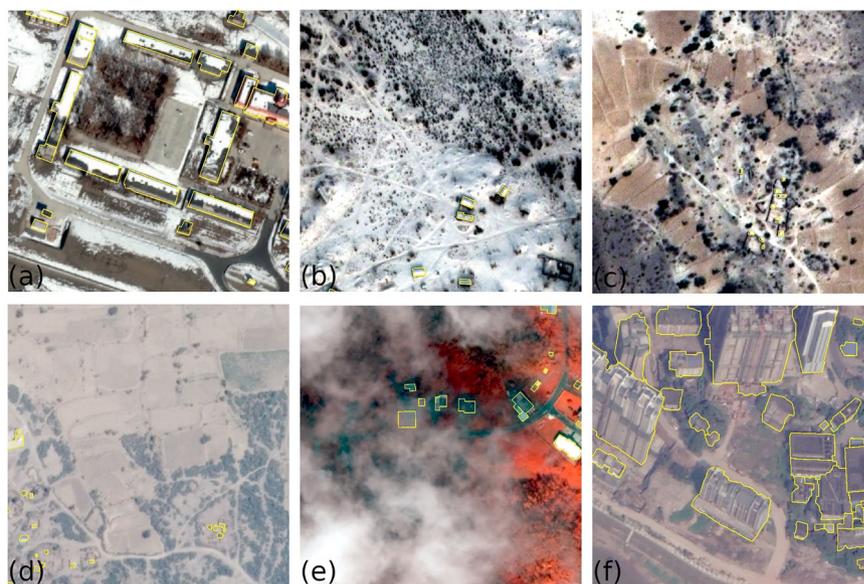


Fig. 11 Mixed roof types (natural material and metal). (a) image, (b) image with labels. ©Maxar 2023.

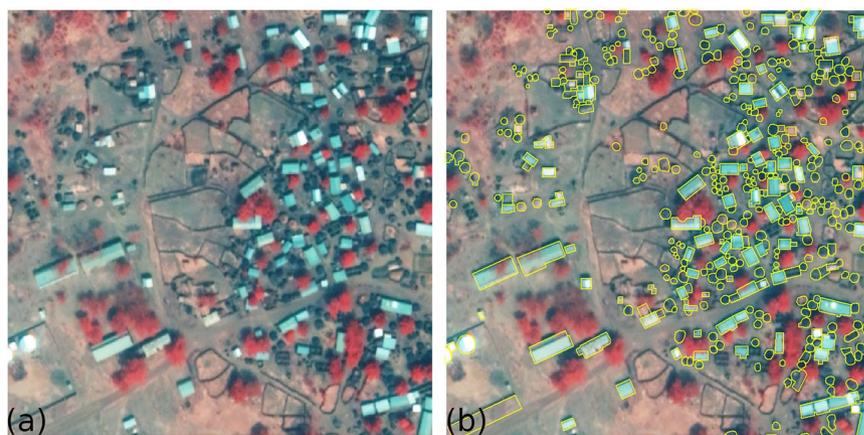


Fig. 12 Training samples included imagery with (a) partially snow-covered ground and buildings, (b) fully snow-covered ground and buildings, (c) grainy imagery, (d) haze, (e) clouds, and (f) extreme off-nadir. ©Maxar 2023.

awnings), and interpret the geographic variation of building types and styles. Furthermore, the regional variation of building styles and types required that analysts had experience interpreting those differences and translating them to labeled training data. For example, walled compounds in Pakistan as opposed to huts in Nigeria, or more broadly how local culture and geography manifests in the built environment. To aide in the task of interpreting regional variations, analysts typically used additional reference imagery to learn the specifics of a region's architecture and settlement patterns and resolve the interpretation of ambiguous features. Examples

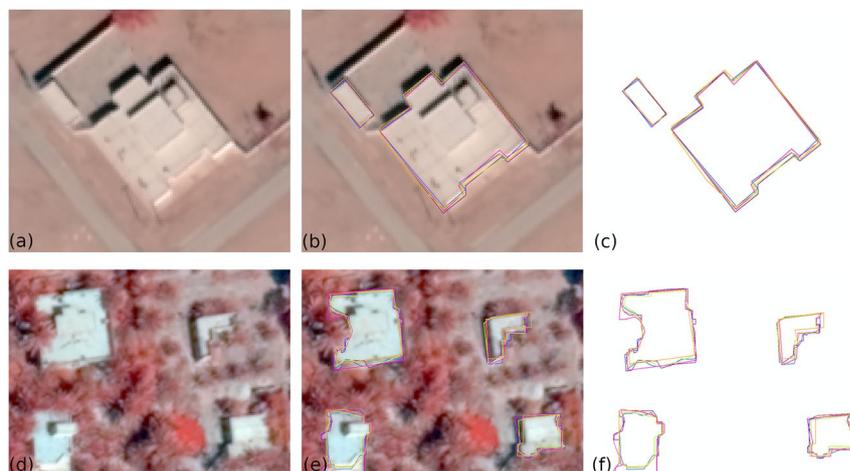


Fig. 13 Examples of acceptable variation in labeled training data developed by five different analysts shown in unique colors. (a,d) show base image. (b,e) show labels on imagery. (c,f) show labels only. ©Maxar 2023.

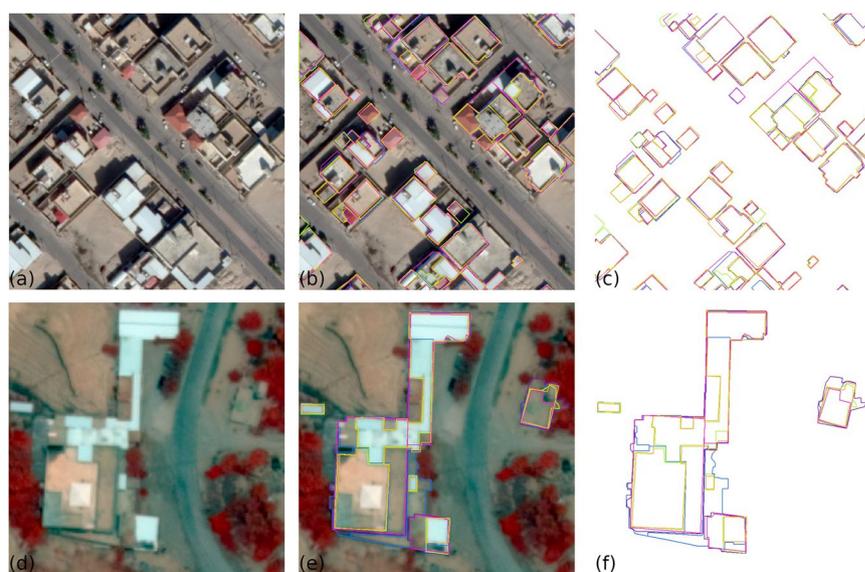


Fig. 14 Examples of unacceptable variation in labeled training data developed by five different analysts shown in unique colors. (a,d) show base image. (b,e) show labels on imagery. (c,f) show labels only. ©Maxar 2023.

include publicly available sources, such as street view imagery from Google, Bing, and Baidu maps, aerial images showing an oblique view, and publications and image libraries of vernacular architecture.

Data Records

ORBITaL-Net consists of 128,270 samples, each comprised of a reference image, a binary training label, and georeferencing data stored in json format, for a total of 384,810 individual files, deposited at Figshare³⁷ <https://doi.org/10.25452/figshare.plus.25282225.v3>. The source imagery dates from between 2003–2020, with the vast majority being from 2010 onwards. The files are in GeoTiff format with the reference imagery held in unsigned 16-bit integer arrays and the labels in binary integer arrays.

The naming convention is as follows:

[source image] [individual image chip ID] [data flag] [sample component]

e.g. 11APR17162634-502717212090_P001-0017_1_img.tif

source image: 11APR17162634-502717212090_P001

individual image chip ID: 0017

data flag: 1

sample component: img

Data flag indicates whether the sample contains building pixels (1) or is purely negative (0)
 Sample component indicates whether the raster is the sample image or one of several types of labels. Values are:
 img: the aerial or satellite image
 seg: binary label of building (1) and non-building (0) pixels

The georeferencing information stored in geojson format is named according the basename of the relevant sample, with- out the “_img” or “_seg” component,
 e.g. “11APR17162634-502717212090_P001_0017_1.guojson”

Technical Validation

To ensure accuracy and consistency across the team of analysts in a variety of challenging contexts, a continuous QA/QC process was implemented. Steps included:

- Analyst self-review
 - The analyst reviewed their work to ensure building features were accurately digitized, no buildings were omitted, and non-building features were excluded from each labeled image tile.
- Buddy-system or expert analyst review
 - A second analyst reviewed labeled image tiles to ensure accuracy and consistency. Edits were made where appropriate.
 - If recurring mistakes were noted, additional instruction was given to the analyst by an expert analyst.
 - Additional examples of frequently misinterpreted/mislabeled features were added to a living document for reference and future instruction.
- Final review
 - Prior to model training, an expert analyst performed a binary review of labeled image tiles. Any labeled tiles that were below standard were omitted from the training dataset. No additional edits were made.

As part of our process of training image analysts, we created a set of training images with expert made “gold-standard” building label that represent a wide variety of imagery and architectural and settlement pattern contexts. By comparing the variation between the expert made labels and those created by a image analyst under training, we could tailor individual instruction. Figures 13 and 14 show examples of acceptable and unacceptable variation, respectively, and, by extension, quality standards which were used for all of the labels in the ORNL dataset.

Statement on authorization for public release. The images included in this dataset were part of MAXAR imagery collection shared with National Geospatial Intelligence Agency (NGA) under NextView agreement. We have obtained the approval (A706) to release this set of images to the public. We also included copyright statement (c 2025 Maxar) into metadata as requested by NGA as part of approval.

Code availability

All pre-processing of the imagery in this dataset was accomplished using open source and freely available algorithms, as detailed in our methods section and in³⁵. No custom code was used to generate these data.

Received: 27 March 2024; Accepted: 24 July 2025;

Published online: 17 October 2025

References

1. Aimaiti, Y., Sanon, C., Koch, M., Baise, L. G. & Moaveni, B. War related building damage assessment in kyiv, ukraine, using sentinel-1 radar and sentinel-2 optical images. *Remote. Sens.* **14**, <https://doi.org/10.3390/rs14246239> (2022).
2. Papatoma-Köhle, M., Schlögl, M. & Fuchs, S. Vulnerability indicators for natural hazards: an innovative selection and weighting approach. *Sci. Reports* **9**, 15026, <https://doi.org/10.1038/s41598-019-50257-2> (2019).
3. Silva, V. *et al.* A Building Classification System for Multi-hazard Risk Assessment. *Int. J. Disaster Risk Sci.* **13**, 161–177, <https://doi.org/10.1007/s13753-022-00400-x> (2022).
4. Evans, G. Accessibility, urban design and the whole journey environment. *Built Environ.* (1978-) **35**, 366–385 (2009).
5. Thrall, G. Gis applications in real estate and related industries. *J. Hous. Res.* **9**, <https://doi.org/10.1080/10835547.1998.12091928> (1998).
6. Rose, A. & Huyck, C. K. Improving catastrophe modeling for business interruption insurance needs. *Risk Analysis* **36**, 1896–1915, <https://doi.org/10.1111/risa.12550> (2016).
7. Allen-Dumas, M. R. *et al.* Impacts of the morphology of new neighborhoods on microclimate and building energy. *Renew. Sustain. Energy Rev.* **133**, 110030, <https://doi.org/10.1016/j.rser.2020.110030> (2020).
8. Wan, K. K. W., Li, D. H. W., Pan, W. & Lam, J. C. Impact of climate change on building energy use in different climate zones and mitigation and adaptation implications. *Appl. Energy* **97**, 274–282, <https://doi.org/10.1016/j.apenergy.2011.11.048> (2012).
9. Deng, M., Menassa, C. & Kamat, V. From bim to digital twins: A systematic review of the evolution of intelligent building representations in the aec-fm industry. *J. Inf. Technol. Constr.* **26**, 58–83, <https://doi.org/10.36680/j.itcon.2021.005> (2021).
10. Ferré-Bigorra, J., Casals, M. & Gangoellés, M. The adoption of urban digital twins. *Cities* **131**, 103905, <https://doi.org/10.1016/j.cities.2022.103905> (2022).
11. Ivanov, S., Nikolskaya, K., Radchenko, G., Sokolinsky, L. & Zymbler, M. Digital twin of city: Concept overview. In *2020 Global Smart Industry Conference (GloSIC)*, 178–186, <https://doi.org/10.1109/GloSIC50886.2020.9267879> (2020).
12. Angjeliu, G., Coronelli, D. & Cardani, G. Development of the simulation model for digital twin applications in historical masonry buildings: The integration between numerical and experimental reality. *Comput. Struct.* **238**, 106282, <https://doi.org/10.1016/j.compstruc.2020.106282> (2020).

13. Cheriyyadat, A., Bright, E., Potere, D. & Bhaduri, B. Mapping of settlements in high-resolution satellite imagery using high performance computing. **69**, 119–129, <https://doi.org/10.1007/s10708-007-9101-0>.
14. Patlolla, D. R., Bright, E. A., Weaver, J. E. & Cheriyyadat, A. M. Accelerating satellite image based large-scale settlement detection with GPU. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, BigSpatial '12, 43–51, <https://doi.org/10.1145/2447481.2447487> (Association for Computing Machinery).
15. Mountrakis, G., Im, J. & Ogole, C. Support vector machines in remote sensing: A review. **66**, 247–259, <https://doi.org/10.1016/j.isprsjprs.2010.11.001>.
16. Camps-Valls, G. & Bruzzone, L. *Kernel Methods for Remote Sensing Data Analysis* (Wiley).
17. Maggiori, E., Tarabalka, Y., Charpiat, G. & Alliez, P. Convolutional neural networks for large-scale remote-sensing image classification. **55**, 645–657, <https://doi.org/10.1109/TGRS.2016.2612821>.
18. Kaiser, P. *et al.* Learning aerial image segmentation from online maps. **55**, 6054–6068, <https://doi.org/10.1109/TGRS.2017.2719738>.
19. Hu, F., Xia, G.-S., Hu, J. & Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. **7**, 14680–14707, <https://doi.org/10.3390/rs71114680>.
20. Yang, H. L. *et al.* Building extraction at scale using convolutional neural network: Mapping of the united states. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **11**, 2600–2614, <https://doi.org/10.1109/JSTARS.2018.2835377> (2018).
21. spacenetai. <https://spacenet.ai/>.
22. ramp. <https://rampml.global/>.
23. Swan, B. *et al.* Orbital-net training library for building extraction. *figshare* <https://doi.org/10.25452/figshare.plus.25282225.v1> (2024).
24. Deep learning for remote sensing data: A technical tutorial on the state of the art. **4**, 22–40, <https://doi.org/10.1109/MGRS.2016.2540798>.
25. Cheng, G., Han, J. & Lu, X. Remote sensing image scene classification: Benchmark and state of the art. **105**, 1865–1883, <https://doi.org/10.1109/JPROC.2017.2675998>.
26. Zhu, X. X. *et al.* Deep learning in remote sensing: A comprehensive review and list of resources. **5**, 8–36, <https://doi.org/10.1109/MGRS.2017.2762307>.
27. Zhou, P. & Feng, J. Understanding generalization and optimization performance of deep CNNs. In Dy, J. & Krause, A. (eds.) *Proceedings of the 35th International Conference on Machine Learning*, vol. 80 of *Proceedings of Machine Learning Research*, 5960–5969 (PMLR, 2018).
28. Christie, G., Fendley, N., Wilson, J. & Mukherjee, R. Functional map of the world. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6172–6180, <https://doi.org/10.1109/CVPR.2018.00646> (IEEE).
29. Challenge, S. <https://spacenet.ai/challenges/>.
30. Sumbul, G., Charfuelan, M., Demir, B. & Markl, V. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 5901–5904, <https://doi.org/10.1109/IGARSS.2019.8900532> (IEEE).
31. xView Challenge. <https://www.diu.mil/ai-xview-challenge>.
32. Gupta, R. *et al.* Creating xbd: A dataset for assessing building damage from satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (2019).
33. Chowdhery, A. *et al.* Palm: Scaling language modeling with pathways. *arxiv:2204.02311* (2022).
34. Osserman, R. The isoperimetric inequality. *Bull. Am. Math. Soc.* **84**, 1182–1238 (1978).
35. Reith, A. *et al.* Providing Geospatial Intelligence through a Scalable Imagery Pipeline. In *Advances in Scalable and Intelligent Geospatial Analytics* Num Pages: 16 (CRC Press, 2023).
36. Swan, B., Laverdiere, M., Yang, D. H. L. & Rose, D. A. Iterative self-organizing scene-level sampling (isosceles) for large-scale building extraction. *GIScience & Remote. Sens.* **59**, 1–16, <https://doi.org/10.1080/15481603.2021.2006433> (2022).
37. Swan, B. *et al.* ORBITaL-Net Training Library for Building Extraction. *Figshare+ Dataset*, <https://doi.org/10.25452/figshare.plus.25282225.v3> (2024).

Acknowledgements

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

Author contributions

B.T.S. co-lead on methods, lead on data analysis, contributed on figure and table creation. J.P. figures and table creation, technical validation. D.R. figures and table creation. H.L.Y. writing on background and motivation, verifying the submitted dataset. A.R. writing, review, contributed to conceptualization and motivation for research. M.L. co-lead on methods.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© UT-Battelle, LLC 2025