Title

Machine Learning for Automated Extraction of Building Geometry

Authors

Touzani, Samir

Pritoni, Marco

Singh, Reshma

et al.

Publication Date

2020-08-01

Peer reviewed

BERKELEY LAB

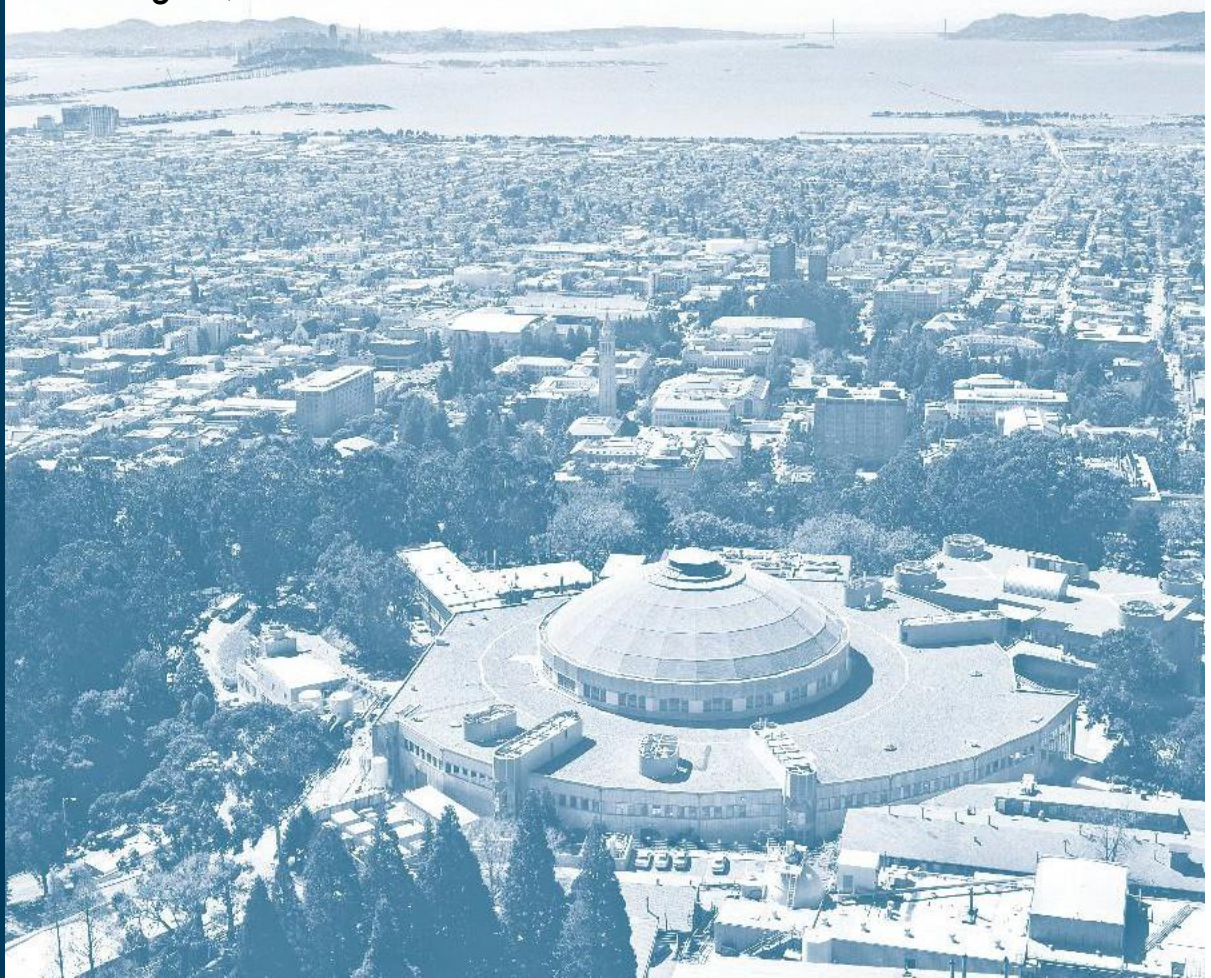
Lawrence Berkeley National Laboratory

Machine Learning for Automated Extraction of Building Geometry

Samir Touzani, Marc Wudunn, Avideh Zakhor, Marco Pritoni, Reshma Singh,
Harry Bergmann and Jessica Granderson

Lawrence Berkeley National Laboratory, Berkeley

Energy Technologies Area
August, 2020



Disclaimer:

Acknowledgments:

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Machine Learning for Automated Extraction of Building Geometry

Samir Touzani[1], Marc Wudunn[2], Avideh Zakhor[2], Marco Pritoni[1], Reshma Singh[1], Harry Bergmann[3] and Jessica Granderson[1]

[1] Lawrence Berkeley National Laboratory
[2] Signetron / UC Berkeley
[3] US Department Of Energy

Abstract
As data science comes to buildings, the promise of using machine learning and novel sources of data has received much attention. Advances in machine learning and computer vision algorithms, combined with increased access to unstructured data (e.g., images and text), have created an opportunity for automated extraction of building characteristics – cost-effectively, and at scale. Acquisition of features such as footprint are time consuming and costly to acquire with today's manual methods, but can be streamlined through intelligent software-based solutions applied to satellite images. When combined with aerial RGB and thermal images, full 3D geometries and thermal maps can be constructed to determine additional characteristics such as window to wall ratio, height, number of stories and envelope thermal characteristics. In this paper we present three contributions to accelerate these high potential opportunities: (1) a methodical analysis of how these features can be integrated into today's simulation and data driven software tools to enhance efficiency measure identification and owner/operator decision making; (2) development and accuracy testing of open source deep neural network methods to extract building footprints from satellite imagery, including the curation and application of openly available GIS datasets for training and continued development by others; and (3) an open framework for drone-based image capture and creation of 3D building geometries. This work represents an important bridge between high-level studies that span diverse application areas and those that detail point solutions yet cannot be easily replicated or extended.

1. Introduction

In recent years advances in the performance of machine learning (ML) methods and computer vision algorithms, combined with increased access to unstructured data (e.g., images and text), have considerably improved the ability to extract information from these types of data. They also have created an opportunity for automated extraction of building characteristics – cost effectively, and at scale. The promise of using ML and unstructured data to improve energy efficiency has received much attention. However, this promise is commonly presented at a high level, without methodical evaluation for specific application areas. One of the primary goals of this project is to provide a concrete assessment of how novel data sources combined with ML can be used to enhance efficiency measure identification in commercial buildings.

State of the art energy efficiency analytics and modeling tools can provide valuable insights into efficiency opportunities. These tools provide a cost-effective and scalable way of identifying energy efficiency measures (Summers et al. 2013). With the ability to enable operational savings on the order of 5 to 10 percent (Granderson and Lin 2016), and to provide retrofit recommendations, they represent a key element in achieving the ambitious goal of reducing energy use per square foot of buildings. However prior research has shown that key barriers include relatively limited data sources (smart meters and weather being most common in commercial tools), or reliance upon user-provided inputs for which default values may be the fallback. The addition of unstructured data, enhanced by ML, have the potential to increase the accuracy and the adoption of these remote analysis and audit tools.

The work presented in this paper aimed at identifying ways in which non-traditional data sources may be used to support identification of energy efficiency measures (EEM) and to improve existing EE tool inputs. In Section 2, we will summarize the performed opportunities analysis to apply ML for enhancing the exiting EE tools. This analysis has suggested that the most promising data sources to fulfill this goal are: satellite and aerial imagery data, drone based visible and thermal imagery data, cities/counties open data, and OpenStreetMap data. Thus, in Section 3 we will describe our exploratory findings on these data sources. From this investigation three data sources surfaced as the most promising in terms of data availability and ease of access. These three data sources will be used in our work for building geometry extraction using ML and computer vision. Section 4 presents the developed ML framework to extract the building footprint from satellite imagery and the computer vision algorithm to extract the building 3D geometry.

2. Analysis of opportunities to apply machine-learning for enhanced energy efficiency

While today's analysis tools are effective and valuable, their outputs can be enhanced even further with contextual information that can be gathered from non-traditional data sources, e.g., age of the building, envelope characteristics, and presence of certain types of equipment. However, the promise of using machine learning and unstructured data to derive this information has not yet been methodically evaluated. We conducted an opportunity analysis that provides a concrete assessment that maps common efficiency measures to potentially relevant data sources, data sources to obtainable features (i.e. building characteristics), and features to the types of tools that they could be used to augment.

Based on considerations of data availability, ease of access, and potential to enhance measure identification, the opportunity analysis suggested that the most promising data sources for this work was: satellite and aerial imagery data, drone based visible and thermal imagery data, city/county open data (e.g., buildings footprints GIS data, building permit data, property taxes data, Lidar data), and OpenStreetMap crowdsourced data. Table 1 connects these four data sources to the feature category identified by the opportunity analysis.

Feature (Category)	Use in Simulation Tools	Use in Data-Driven Tools	Data Sources
Devices/Equipment Inventory (Equipment)	Used to refine assumptions related to equipment (type, efficiency level)	Used to refine assumptions related to equipment installed, exclude EEM (e.g., skylights already present), and quantify cost in the cost-benefit analysis (i.e., cost can be related to the number of units to be replaced)	Aerial imagery Drone imagery Building permits
Building Footprint (Envelope)	Used to calibrate auto-generated building models	Can be used in combination with building external geometry to normalize energy savings for benchmarking with other buildings	Building footprint data Aerial imagery Satellite imagery Drone imagery OpenStreetMap
Building Vintage (Envelope)	Used to refine the building vintage assumptions	Used to infer the most likely type of envelope and equipment	Property taxes data OpenStreetMap
Building External Geometry (Envelope)	Used to refine inputs related to the building envelope	Used to identify the building, # of floors, and shading that may prioritize certain EEMs	Aerial imagery Satellite imagery Drone imagery
Building Envelope Characteristics (Envelope)	Used to refine inputs related to the building envelope	Used to assess the need for envelope measures, such as adding insulation or reducing leakage	Aerial imagery Drone imagery
Shading (Surroundings)	Used to refine geometry modeling of building block simulations	Used to evaluate the potential of using shading as an EEM	Aerial imagery Drone imagery
Building/Business Opening Hours (Occupancy)	Used to refine inputs related to operating hours or number of occupants, and schedules	Used in analysis to identify EEMs that impact scheduling of equipment and operation	OpenStreetMap
Building Use (Context)	Used to select a reference model	Used to exclude EEMs related to non-pertinent systems (e.g., process load in an office) or to have a better estimate of a sub-system load	OpenStreetMap Property taxes data
Building Change Records (Context)	Used to define equipment characteristics	Used to identify whether systems have recently been updated or replaced. May exclude some EEMs	Building permit data

Table 1. Mapping features to EE tools and data sources

3. Unstructured Data sources

Satellite and aerial imagery Data

High resolution satellite and aerial imagery, which provide a vertical view of the earth, have been increasingly popular and widely used in several applications. Several recent studies (Demir et al. 2018; Gavankar and Ghosh 2018) have explored the use of these two types of imagery for automatic extraction of building footprints. The main difference (that is relevant to our work) between aerial and satellite images is the resolution. While most of the recent observation satellites can capture images of up to 30 centimeter (cm) resolution (i.e., 1 pixel = 30 cm), it is possible to capture aerial images with a resolution of up to 7 cm. The higher resolution of the aerial images can also provide more information about a building's rooftop equipment (e.g., detection of a cooling tower and/or packaged rooftop unit). Another advantage of the aerial imagery is that it can be captured as oblique imagery (i.e., captured at approximately 45° with the ground), which allows users to analyze building façades. The oblique images can be used to detect/segment façade features (e.g., windows, doors) and estimate some of the building exterior characteristics (e.g., height, window to wall ratio).

Several providers offer high resolution satellite or aerial images, either through an application programming interface (API) or a direct download. However, one of the biggest challenges of using data that can be collected from these imagery companies is the licensing agreement, which can tightly limit the type of analysis that we can perform on these datasets and on the eventuality of sharing the derivate works (e.g., extracting a footprint from aerial images). Several providers also allow free access to satellite and aerial imagery (e.g., Bing map, Google Maps, Mapbox), but they do not grant the user the right to download the images and to create derivative works based on them. Another potential source for gathering free aerial vertical imagery is cities' open data portals. Cities such as New York City, Los Angeles, Portland, and Washington D.C. are regularly releasing high resolution aerial imagery through their open data platform. Note that to the best of our knowledge the only freely available aerial oblique imagery is the National Oceanic and Atmospheric Administration (NOAA) coastal oblique imagery, however this dataset is limited to the coastal cities of the United States.

(Cities) Open Data

An increasing number of U.S. cities are providing different public building datasets through an easy to access open data web portal. The three most relevant data sources that can provide relevant information about building characteristics are building footprint data, building permit data, and property tax (Assessor record) data.

Building footprint data

Building footprints are provided by several cities (e.g., San Francisco, Washington D.C., Boston, Los Angeles, Chicago, New York) through an open data portal. It is usually accessible via an API or direct download, and it is available as a geographic information system (GIS) data

file. In some cases, the height and some type of building ID is also provided. Using some preexisting GIS tools, this type of dataset is generally easy to process. Note that although most of the explored cities regularly update these datasets, some cities only provide building footprint information that has not been updated for the last 5 to 10 years.

Building permit data

Building permit data also is provided by several cities (e.g., San Francisco, Washington D.C., Boston, Los Angeles, Chicago, New York) through their open data portals. Accessible via an API or direct download, it comes as tabular data, mostly as JSON or CSV files. Each permit is usually georeferenced either by the postal address and/or by the geolocation (i.e., latitude and longitude). This type of dataset is the most challenging to process because they lack of uniformity and semantic standardization in the permits description. In addition, some cities provide a direct access for several years of permits in one unique dataset, while others will store the information in multiple files, sometimes changing the semantic in the description and the type of information that is provided from one year to another. The buildings features that can be extracted from this type of datasets are: building change records and building use.

Property tax (Assessor record) data

Several cities (e.g., San Francisco, New York, Boston and Baltimore) provide this type of dataset openly, but it is more challenging to obtain in comparison to the previous two datasets. However, when available, it provides information on several building features such as: floor area, vintage, number of stories, number of dwellings, and type of use. As with building permit data, this type of dataset is shared as tabular data, where each building is georeferenced using the postal address and/or by its geolocation (i.e., latitude and longitude).

OpenStreetMap

OpenStreetMap (OSM) is currently the largest openly licensed database of geospatial data (i.e., GIS data). It is a global dataset that is a collaborative product (i.e., crowd sourced data) created mainly by more than one million volunteers who have contributed to editing the database content (OSM Wiki: Stats 2019). It includes different types of GIS information, such as road infrastructure and built environment, and it is used in many projects as an alternative to proprietary or authoritative data. While the crowd sourced aspect of OSM has been a key to its success, the low number of volunteers with professional GIS experience has raised significant concerns about its accuracy. Several studies (Schiefelbein et al. 2019; Brovelli and Zamboni 2018) have performed an assessment of OSM spatial accuracy and completeness of buildings footprints. However, because all these studies have focused their analysis on very limited geographical regions it is not reliable enough to extrapolate a general conclusion regarding the quality of OSM in terms of building footprint.

We conducted on a sample of U.S. cities a qualitative analysis of the OSM data accuracy and availability regarding building footprints and other building characteristics relevant to this

project. Our findings show that the quality and availability vary significantly, depending on the city considered, and that this variability is highly correlated to whether or not the city's authorities have made the building footprint openly available. For example, the quality, and more specifically the completeness, of the OSM data for the City of Las Vegas is very low, and to the best of our knowledge the city's authorities do not openly provide complete building footprint GIS data. Figure 1 shows an example of a comparison between the OSM map and the satellite image (captured from a Google map) of specific region in Las Vegas. The OSM data has a very limited building footprint information.

Figure 1. Left: an OSM map from a specific region in Las Vegas; Right: a satellite image from the same image taken from Google map.

Note that, in addition to the building footprint, the OSM database can provide additional information regarding building characteristics, such as: building height, type of building use, operation hours for some commercial buildings, and building vintage. However, our analysis has shown that only a very limited number of buildings have this information available.

Drone imagery data

In the past decade, the rapid advances in drone technologies (e.g., aeronautics, sensing, autonomous navigation systems) have made these systems more reliable, easier to operate, and significantly cheaper, which has resulted in an exponential growth of their use for monitoring building construction and operation. A drone system can easily and quickly capture a large number of images of buildings from different angles, and particularly can facilitate the inspection of areas that are hard to reach without compromising operator safety. Combined with an RGB camera and a thermal camera, as well as with advanced computer vision tools, the drone system presents an opportunity to streamline the building energy auditing process by analyzing buildings' envelopes efficiently and accurately at a reduced operational cost in comparison to a more traditional energy auditing process (Rakha and Gorodetsky 2018). In comparison to 3D building geometry information that is available through city and county open data portals, or that can be extracted using satellite/aerial imagery or Lidar data, drone-based approaches can produce higher resolution information, enabling more accurate estimation of building geometrical characteristics (e.g., footprint, height, window to wall ratio). In addition, the use of thermal cameras can provide insights into the building envelope that none of the other data sources can offer.

To define the specifications of the drone system that we have used in our data acquisition, the following list of criteria has been used:

- Type of sites where data will be acquired: individual buildings and clusters of buildings (e.g., a block of buildings or a neighborhood)
- Type of information that will be extracted: 3D building model generation, façade and roof segmentation (e.g., detection of windows, doors) and thermal leakages detection

The selected drone system is composed from a quadcopter drone, which has some advantages in comparison to other drone types (e.g., fixed wing, helicopter). The main advantages are that a quadcopter can hover and is able to fly very slowly, which are two characteristic that can be important for obtaining high resolution images, especially from building façades. Other advantages of this type of drone are the low cost in terms of purchase price, operation and maintenance; very flexible operation, especially in cities (i.e., they do not require a large area to take off and land); and relative ease to control in autonomous and manual pilot modes. The selected drone also has dual gimbals that enable it to operate the RGB and the thermal camera simultaneously. The RGB camera is a high-resolution camera (20.8 megapixels) that can be used in still (capturing static pictures) or video mode. This system has also been paired with a high resolution (640 × 512) Radiometric thermal camera that has an infrared sensor, which can spot pixel temperature with relatively good precision (Al Lafi 2017).

4. Feature extraction using machine learning and computer vision approaches

The previous section presented the investigation, which was based on the availability, quality and potential impact of the features that can be extracted from different data sources that emerged from the opportunity analysis. Nevertheless, three sources surfaced as most promising, and those were used in our work for feature extraction using machine learning and computer vision. These three data sources are (1) satellite/aerial images, (2) building footprint GIS data and (3) drone-based visible and thermal images. Their strengths are shown below.

- **Satellite/aerial images**
  - These images are highly available because several providers can provide access to images that cover most major U.S. cities.
  - They offer a high potential to scale because of the high coverage of these data sources and the potential to automate feature extraction via ML and computer vision algorithms.
  - Several sources of openly available data exist that can be used for R&D.
- **Building footprint GIS data**
  - These data are openly available for many U.S. cities and are relatively easy to extract and process.
  - Combined with satellite/aerial images and ML, this source can provide a large

labeled dataset that can be leveraged to build accurate models to extract building footprints of U.S. regions where GIS footprint data are not available.

- **Drone-based visible and thermal images**
  - These high-resolution images (up to 10 times higher than satellite images) can be captured easily, quickly and relatively cost-effectively.
  - They offer a high potential to use thermal data to assess building envelope characteristics.
  - These images can facilitate inspection of areas that are hard to reach, without compromising safety.

In the following sections we will present the ongoing work on the building geometrical features extraction from the selected data sources using ML and computer vision approaches.

4.1 ML-based satellite/aerial image feature extraction

In recent years advances in the performance of ML methods have considerably improved the ability to accurately extract information from images. Identification of building footprints from satellite/aerial imagery is an example of an application that is time consuming and costly to perform manually, and where automatic feature extraction holds great promise. This application has therefore triggered several attempts to develop ML based solutions. These prior works (USBuildingFootprints 2018, Van Etten et al. 2018; Maggiori et al. 2017), however, are limited in that the data and models are either not available for use by others or are too limited for development of generalized solutions. A major contribution of this project is the provision of an open and replicable solution for building feature extraction from satellite/aerial images. The majority of these prior research efforts have framed the challenge as a semantic segmentation problem, which is a computer vision task that partitions the image into semantically meaningful predefined classes. In other words, each pixel of the image is assigned a class.

Recently, Microsoft (USBuildingFootprints 2018) published building footprints for all buildings in the United States. These footprints were generated from aerial/satellite images using a deep learning based semantic segmentation approach. While the Microsoft dataset provides the building footprint GIS data, it does not make available the images that were used to train the model. Moreover, only a limited description of the model and the training process is provided. Since the quality of the footprint information is somewhat limited, and the data represents a snapshot in time of the building stock, it is useful (but not possible with the information Microsoft has released) to be able to both replicate and improve upon the current footprint inventory. In addition to the Microsoft work, several benchmark datasets include both aerial/satellite images and building footprints (manually labeled) (Van Etten et al. 2018; Maggiori et al. 2017). However, these datasets are limited to a small number of U.S. cities, making it difficult to use them to develop a model that can be generalized to the entire U.S. territory.

In this project, we initiated the development of a replicable workflow to extract building footprints from satellite/aerial images using a state of the art deep learning image segmentation algorithm. This workflow is composed of three modules:

- A data preparation module, which aims to generate training data with a very limited manual effort using openly available data sources:
    - Automatic generation of training features masks by querying city/county footprint GIS open data
    - Automatic extraction of satellite/aerial tiles using Mapbox. These tiles cover the geographical areas that are available from the city/county footprint GIS open data
- Deep learning modeling, which aims to establish an easy modeling pipeline to reproduce the analysis and test new deep learning architectures:
    - Standardized training pipeline
    - A state of the art fully convolutional neural network architecture models (i.e., DeeplabV3+)
- A postprocessing of model results module, which aims to generate results that are easily transformable into data formats that are compatible with required inputs to existing measure-identification tools:
    - Prediction cleaning (i.e., removal of noisy prediction)
    - Prediction transformation (i.e., convert predictions that are pixel-based masks into polygons with geographic coordinates)

Application of the workflow

Data capture and preparation

To date, we have collected building footprint GIS data from 14 cities and 6 counties, with the total dataset covering 8,607,677 buildings. For each GIS building footprint dataset collected, we created a list of tile coordinates that cover the specific region. These coordinates follow the Mapbox Raster Tiles API default format, which is the slippy maps (OSM Wiki: Slippy map 2020) standard that defines each tile by the zoom level, which defines the resolution and the tile coordinates. Satellite images were collected from the satellite layer of the Mapbox Maps (Mapbox 2019), which provides high resolution satellite images. This imagery data is a compilation of several commercial and open data sources, and is color-corrected and stored in raster format, which is a pixel-based data format "that efficiently represent continuous surfaces" (Mapbox 2019). These satellite images can be retrieved relatively easily using the Mapbox Raster Tiles API. The zoom level that was considered in this work was 19. This level was chosen based on an empirical analysis that tested several zoom levels in the machine learning pipeline and compared the accuracy of the results on a very limited dataset. The total number of images collected at zoom level 19 that cover geographical regions for which building footprints are

available is 2,432,019. The corresponding feature mask images, which are the representation of the buildings footprints that are contained in the corresponding satellite images, were produced by reformatting the GIS files into binary images. Note that the feature mask images have the same resolution as the satellite image tiles. To avoid overlapping masks between adjacent buildings, we decreased the size of the footprint polygons by a factor of ~8%, in terms of surface area.

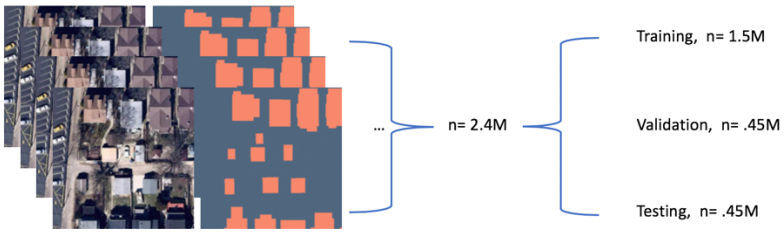


Figure 2. Illustration of the data preparation process

The full dataset is divided into three samples to be used for training, validation and testing (see Figure 2). Each sample contains pairs of a feature mask image and a satellite image that were randomly selected. The training dataset has ~60 percent of the total number of pairs (n = 1.5M pairs), the validation dataset has ~ 20 percent (n = .45M pairs) and the test dataset also has ~20 percent (n = .45M).

Deep learning model tuning and training

In this step of the process, deep learning model hyperparameters are tuned. Data from the training sample are used to fit the models using different combination of hyperparameters. The models are then used to predict footprints from the validation sample. The model-hyperparameter combination(s) that exhibit the best performance can then be evaluated for overall accuracy using data from the testing sample. In the general case supported in this workflow, any model can be explored. In the work conducted specifically for this project, we have explored deep convolutional neural networks that are deployed in fully convolutional fashion – that is, the neural networks are composed *only* of convolutional layers and do not contain fully connected layers (i.e., all the output neurons are connected to all input neurons) just before the output. This type of neural network architecture has been shown to be very effective for semantic segmentation because they tend to learn spatial information.

We experimented with state of the art fully convolutional neural network model Deeplab V3+ (Chen et al. 2018). Given computational run-time constraints, a limited portion (i.e., ~5 %) of the training and validation samples were used for model tuning. The hyperparameters considered were: optimization algorithms, learning rate, learning rate decay strategies, loss functions, and data augmentation strategies. Even with the high-performance graphics processing unit (GPU) that was used, each assessment (i.e., one configuration of hyperparameters) took several days of computation. The top performing combination of hyperparameters (see AutoBFE

2019) was used to train a model on the full dataset using multi-GPU high performance computers.

Postprocessing the predictions

The final step in the workflow is the postprocessing. The prediction masks that are produced using the neural network model need to be transformed to become meaningful footprint features that can be input into existing measure identification analysis tools. This requires the use of image processing algorithms such as morphological operation (to remove noise in the prediction mask), contouring, and polygonization (to generate GIS polygons from the pixel-based mask).

Accuracy assessment process

The accuracy of the trained Deeplab V3+ model is measured using the F1 score (F1), the precision and the recall. These metrics are based on pixel predicted values represented by the total number of true positive (TP), false positive (FP) and false negative (FN), which are computed using the number of pixels classified as buildings in the predicted masks and in the ground truth masks. Thus:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = \frac{2TP}{2TP + FP + FN} = 2\frac{precision \times recall}{precision + recall} \quad (3)$$

To assess the accuracy, we performed two analyses. The first one used the test sample of ~.45M satellite images that was previously generated by randomly selecting 20 percent of the whole dataset. The results of this analysis are shown in Table 2.

Table 2. **Summary of accuracy results using the test sample**

	Precision	Recall	F1 score
Test sample	0.91	0.89	.90

The second analysis was performed by using the trained Deeplab V3+ model to make a prediction for a small region of Manhattan in New York City (i.e., a very dense urban area of about 8,000 buildings). This analysis was performed to assess the quality of the prediction in comparison to the footprints recently released by Microsoft. Table 3 summarizes the accuracy metrics computed using our predicted footprints and the footprints provided by Microsoft. Note that the metrics have been computed using as ground truth the footprints gathered from NY city open data web portal. For this particular test our predictions, in comparison to the predictions provided by Microsoft, produced better results in terms of Recall and F1 score, which means our

approach had a higher rate of correctly classifying pixels that correspond to a building footprint. Note that the higher precision metric for Microsoft data means our approach has a higher rate of false positives among the pixels that were classified as buildings.

Table 3. Summary of accuracy results

	Precision	Recall	F1 score
Trained Deeplab V3+	0.89	0.96	0.92
Microsoft	0.93	0.83	0.88

Figure 3 shows two examples of the difference in prediction between our approach and the footprints provided by Microsoft. The Microsoft footprint did not detect two adjacent buildings as being separate, while our approach performed better in this task. However, this is not always the case, as shown in example 2. In many situations our model also failed to find the separation between two adjacent buildings, and this can be due to many reasons (e.g., complex roof structure, shading, vegetation).

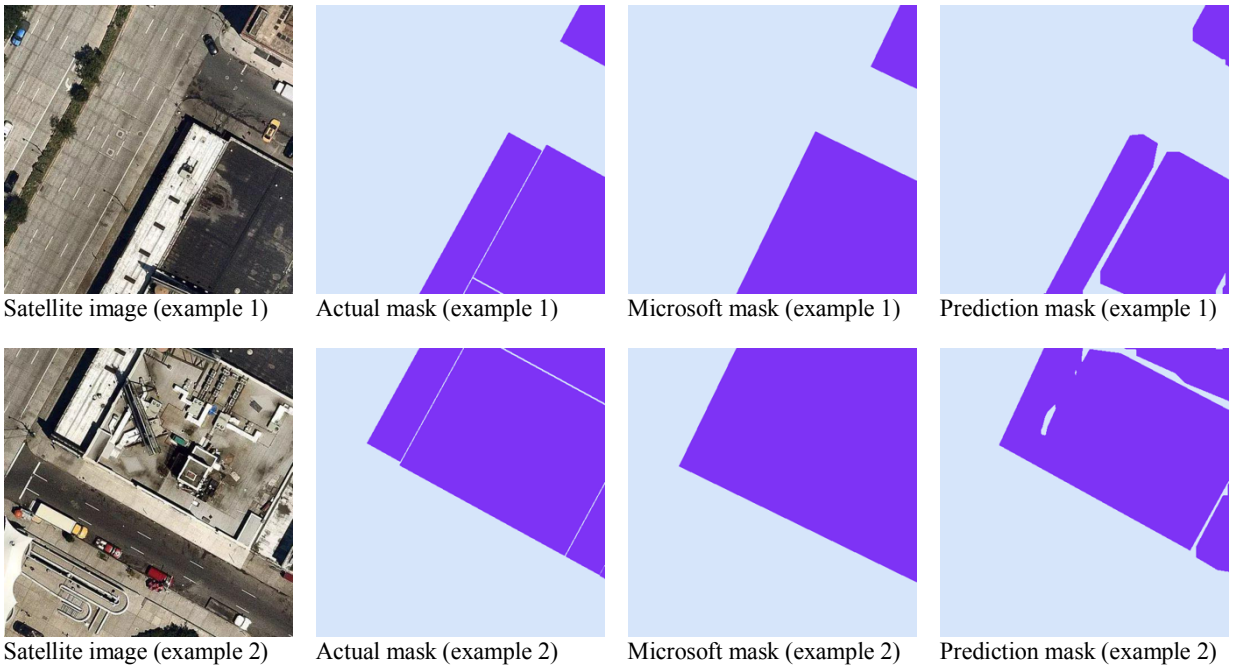


Figure 3. Two examples of prediction results with a comparison to Microsoft building footprints dataset and the ground truth

4.2 Computer vision approach to extract building 3D geometry from drone images

In this section we present the drone-based data capture process and the method developed to extract building 3D geometry from drone RGB images (for more details see WuDunn et al. 2020). To describe this, a case study of an abandoned Naval Air Station office building is used. This building is located in Alameda, California.

Flight Procedure

The drone flight was planned so the drone would fly roughly 20 to 30 feet above the building, with a camera angle of 45◦ below the horizontal. To facilitate the 3D reconstruction of a building, each of its parts must appear in several images taken at different angles. Therefore, images were captured so that successive images have a 90 percent side overlap. The drone flight paths were a series of overlapping circles, as shown in Figure 4. This ensures that images taken are sufficiently close to the building, to allow a good reconstruction.

Photogrammetry

The set of overlapping 2D RGB images captured by the drone system are processed using a photogrammetry software (i.e., Pix4D), which generates a set of data points in 3D space (Figure 5. This is achieved by recognizing the image content and finding tie points between images. This set of data points in a 3D space is called a *point cloud.* Pix4D (Pix4D 2019) uses a fully automated process to achieve accurate 3D reconstruction based on the 2D image sets.

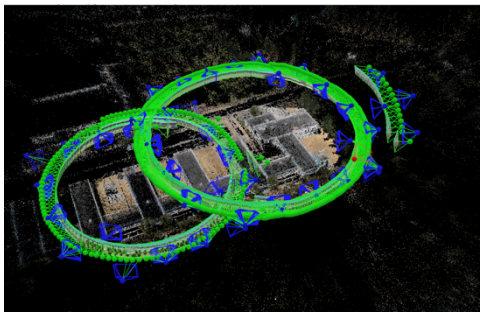


Figure 4. Position of the drone during the data capture

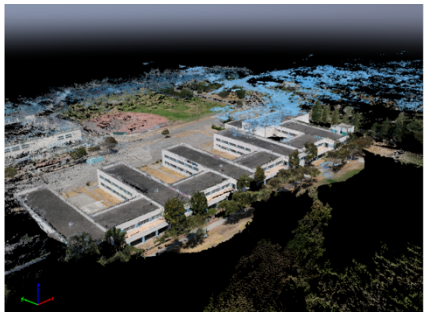


Figure 5. 3D point clouds

Projection of the 3D point clouds into a 2D space

After the point cloud is extracted, the points in the cloud are projected into a 2D grid (top-down view), with a resolution of 0.1meter (m). This permits a count of the number of points in a 0.1m x 0.1m area. This can be visualized in Figure 6, with the white areas indicating a cell with a higher point count. Note that the walls of the building are clearly visible in the image, as each wall has a very high point density when viewed from the top down.

Figure 6. 2D projection of the 3D point cloud

Line Detection and poligonization

We apply the Hough Transform to the resulting grid (i.e., 2D projection). This is a well-known method for detecting lines in an image, by converting an image to “Hough space.” First

note that a line can be represented by two parameters in 2D space: an *angle* (rotation) of a horizontal line passing through the origin, and an *offset* from the origin. A Hough space is an *N*x*M* matrix where each entry represents a possible line. The rows represent all possible distances of a line from the origin, and the columns represent all possible angles (from -180° to 180°). The scalar value of the entries represents the total number of points that the line passes through. By applying the Hough Transform, and finding peaks in the resulting matrix, we can extract lines in the 2D projection. We further refine the lines by applying a developed algorithm that processes the lines and extracts the segments that likely correspond to walls (Figure 7). Once the line segments have been extracted, polygons are constructed. As the line extraction process has various limitations and likely either misses a wall or does not extract the full line segment corresponding to a wall, we have developed an algorithm that completes the polygons by taking a set of edges that would *nearly* form a polygon, and fill in the gaps (Figure 8). Note that the fifth wing on the left of Figure 7 and Figure 8 was not considered during the polygonization step because of the missing walls in the point cloud, which is due to the fact that we only targeted the four right wings of the building when we defined our flight path.

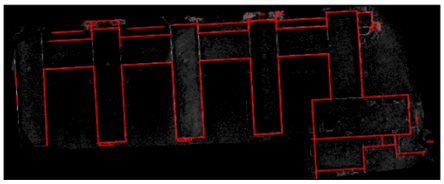


Figure 7. Lines segments that are likely walls

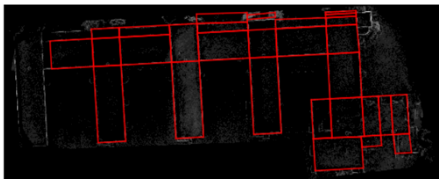


Figure 8. Resulting polygonization

Merging polygons and height estimation

Once all the polygons have been completed, we recursively merge adjacent polygons if they are close together in (average) height. To compute the average height of the polygon, we compute the average height of the points (from the 3D points cloud) in each cell, and then average the heights of the cells. This essentially "spreads out" the height computation across the full area of the polygon, minimizing the effects of objects that may lie on top of the roof. We consider two polygons to be close together if they lie within 0.5 meters of each other. To do so, we can simply iterate through each edge that is used in two polygons and check their heights. After merging two polygons, we update the height of the merged polygon by simply performing a weighted average of the average heights of the two polygons, where the weighting factor is just the number of cells that lie within a polygon. The merged polygons are shown in Figure 9, and the resulting heatmap of the heights of the merged polygons is shown in Figure 10.

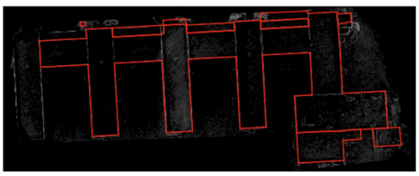


Figure 9. Resulting polygons after the merge process

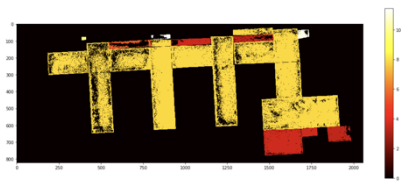


Figure 10. Heatmap of the heights of merged polygons

5. Conclusion and future work

In our initial work we selected three data sources that were determined to be the most promising sources in term of data availability, quality and the potential uses of the features that can be extracted. Using the satellite imagery and openly available footprint GIS data, we developed and applied an ML workflow to extract the building footprint. By developing this workflow and making it openly available (AutoBFE 2019), we are providing a toolbox that allows others to easily reproduce our approach, that offers transparency in modeling solutions, and that facilitates continuous improvement via incorporation of alternative deep learning algorithms and/or expansion of the training data (e.g., adding more cities into the training dataset). Our preliminary analysis using data gathered from 14 cities and 6 counties have shown promising results in comparison, with existing work (i.e., Microsoft's building footprint results). We also developed a computer vision-based building footprint extraction algorithm that uses as an input points cloud generated by photogrammetry from drone imagery. This algorithm also provides an estimation of buildings heights, which enable users to extract the 3D geometry of the building.

Building upon the work presented in this paper, we are improving the postprocessing module of the building footprint extraction ML tool (i.e., tuning some postprocessing hyperparameters). We will also explore the potential improvement of the model accuracy by using higher resolution aerial data combined with Lidar data. In addition, we are developing a machine learning method to estimate window to wall ratio using drone imagery, as well as designing a pipeline to capture thermal imagery and extract thermal features.

Acknowledgment

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Bibliography

Al Lafi, G., 2017. *3D Thermal Modeling of Built Environments Using Visual and Infrared Sensing* (Doctoral dissertation, Concordia University).

AutoBFE 2019. https://github.com/LBNL-ETA/AutoBFE

Brovelli, M. and Zamboni, G., 2018. A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. *ISPRS International Journal of Geo-Information*, *7*(8), p.289.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801–818).

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D. and Raska, R., 2018, June. Deepglobe 2018: A challenge to parse the earth through satellite images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 172–17209). IEEE.

Gavankar, N.L. and Ghosh, S.K., 2018. Automatic building footprint extraction from high-resolution satellite image using mathematical morphology. *European Journal of Remote Sensing*, *51*(1), pp.182–193.

Granderson, J. and Lin, G. 2016. Building energy information systems: Synthesis of costs, savings, and best-practice uses. *Energy Efficiency* 9(6): 1369–1384.

Maggiori, E., Tarabalka, Y., Charpiat, G. and Alliez, P., 2017, July. Can semantic labeling methods generalize to any city? Th Inria aerial image labeling benchmark. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3226–3229). IEEE.

Mapbox, 2019. Retrieved September 25, 2019 from https://docs.mapbox.com/

OSM Wiki: Stats 2019. Retrieved May 21, 2019, https://wiki.openstreetmap.org/wiki/Stats

OSM Wiki: Slippy map 2020. https://wiki.openstreetmap.org/wiki/Slippy_Map

Pix4D, 2019. Pix4Dmapper. Retrieved May 20, 2019, from https://www.pix4d.com

Rakha, T. and Gorodetsky, A., 2018. Review of Unmanned Aerial System (UAS) applications in the built environment: Towards automated building inspection procedures using drones. *Automation in Construction*, *93*, pp.252–264.

Schiefelbein, J., Rudnick, J., Scholl, A., Remmen, P., Fuchs, M. and Müller, D., 2019. Automated urban energy system modeling and thermal building simulation based on OpenStreetMap data sets. *Building and Environment*, *149*, pp.630–639.

Summers, H., et al. FirstFuel Scaled Field Placement. Prepared for Pacific Gas & Electric Co., February 2013.

USBuildingFootprints, 2018. Microsoft, https://github.com/microsoft/USBuildingFootprints

Van Etten, A., Lindenbaum, D. and Bacastow, T.M., 2018. Spacenet: A remote sensing dataset and challenge series. *arXiv preprint arXiv:1807.01232*.

WuDunn, M., Zakhor, A., Touzani, S. and Granderson, J., 2020, June. Aerial 3D building reconstruction from RGB drone imagery. In *Geospatial Informatics X* (Vol. 11398, p. 1139803). International Society for Optics and Photonics.