Applied Energy 280 (2020) 116018

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Estimating residential building energy consumption using overhead imagery

Artem Streltsov [a,*], Jordan M. Malof [b], Bohao Huang [b], Kyle Bradbury [a,b]

[a] Energy Initiative, Duke University, Durham, NC, USA
[b] Department of Electrical & Computer Engineering, Duke University, Durham, NC, USA

HIGHLIGHTS

- Building energy consumption is predicted from overhead imagery alone.
- Buildings are detected and classified by type using deep learning.
- Predict individual building energy consumption explaining 28% of variance.
- Small amounts of spatial aggregation explain 91% of variance in predicted energy.
- Model demonstrated at a practical scale in two locations in the United States.

ARTICLE INFO

Keywords:
Buildings
Energy demand
Energy consumption
Convolutional neural network
Random forest

ABSTRACT

Residential buildings account for a large proportion of global energy consumption in both low- and high- income countries. Efficient planning to meet building energy needs while increasing operational, economic, and environmental efficiency requires accurate, high spatial resolution information on energy consumption. Such information is difficult to acquire and most models for estimating residential building energy consumption require detailed knowledge of individual homes and communities which are unlikely to be available at a large scale.

To address this need, we introduce a methodology for automatically estimating individual building energy consumption from overhead imagery (e.g. satellite, aerial) and demonstrate the effect of spatial aggregation for further improving accuracy. We use a three-step estimation process by which we (1) automatically segment buildings in overhead imagery using a convolutional neural network and classify them by type (residential or commercial), (2) extract features (e.g. area, perimeter, building density) from those identified residential buildings, and (3) use random forests regression to estimate building energy consumption from those features.

The predictive capability of this approach is evaluated in two locations: Gainesville, Florida, and San Diego, California. The building detector correctly identifies 84% and 88% of buildings in Gainesville and San Diego, respectively. The type of building is classified successfully 99% of the time for residential buildings and 74% of the time for commercial buildings. With residential buildings identified, this approach predicted individual building-level energy consumption with an R^2 of 0.28 and 0.38 for Gainesville and San Diego, respectively. Aggregating the energy consumption estimates across small neighborhoods of size 200 × 200 m and 1000 × 1000 m in Gainesville results in an R^2 of 0.91 and 0.97, respectively. We also explore the sensitivity of estimates in San Diego and Gainesville to the training data and its size. Our results suggest that using overhead imagery to estimate the size of buildings has a higher predictive power in estimating residential building energy consumption than common alternatives.

* Corresponding author.
E-mail address: artem.streltsov@duke.edu (A. Streltsov).

https://doi.org/10.1016/j.apenergy.2020.116018
Received 16 February 2020; Received in revised form 25 September 2020; Accepted 9 October 2020
Available online 21 October 2020

1. Introduction

1.1. The need for residential building energy consumption information

Buildings account for more than one third of global energy consumption [1]. In the United States and the European Union, buildings account for more than 40% of energy consumption [2]. In the United States, buildings constitute more than 72% of electricity consumption [3]. Behind these figures is an unprecedented, decades-long trend toward urbanization driven by both rapid population growth and a migration from rural settings to cities.

Urbanization has driven economic development (and vice versa) but these twin trends also result in greater energy consumption and greenhouse gas emissions. In 2008, only half the world's population lived in urban areas, but they were responsible for two-thirds of the world's energy consumption. By 2030, nearly three-fourths of the world's population is projected to live in urban areas [3]. At that point, the United Nations estimates the number of city-dwellers worldwide will grow at a rate of two million people per week. That astounding growth can be managed with careful planning, but if urbanization continues on a mostly ad hoc basis, environmental and social consequences could be severe [4].

While national and international policies will have a role to play in addressing greenhouse gas emissions, cities—where sustainable infrastructure and policy measures find their most immediate implementation—may be in the best position to curb local emissions and energy consumption, including that of buildings.

However, for policymakers to act meaningfully to curb energy consumption, including building-level energy consumption, they first need accurate, actionable data. Detailed building-level energy-use data has many policy and technological applications. For example, researchers have recently investigated the relationship between energy use (e.g., building consumption), economic indicators [5], and climate change [6]; architects and engineers develop 'zero-energy building' technologies aspiring to bring us closer to sustainable growth [2]; and advanced energy use forecasting models offer electricity producers and manufacturers opportunities for cost savings and prudent investments in energy infrastructure.

Individual utility companies have billing data that contains measurements of building energy consumption. Those data are largely unavailable to the research and policy community and typically the only option for accessing them is through non-disclosure agreements, and those are often rare. The California Public Utilities Commission, for instance, has a mechanism for disclosure of energy usage data for approved research, but discloser of the data requires that the individual customer energy data be aggregated to a level with at least 100 consumers per observation [7].

Energy data that are available often have their limitations. The largest of which are either location-anonymized, such as the Residential Energy Consumption Survey (RECS) [8] or the Pecan Street Dataport dataset [9]. Others are limited in coverage area such as Gainesville Green which only includes Gainesville, Florida, [10] or one of the many non-intrusive load monitoring (NILM) datasets (REDD [11], BLUED [12], PLAID [13], GREEND [14], AMPds [15], Tracebase [16], UKDale [17]) that only include a handful of homes each. For addressing research and policy questions in a specific neighborhood or region, these existing datasets, alone, will not provide adequate information, but need to be supplemented with information specific to the study region.

Estimation of energy consumption is one way to overcome these data access limitations. Current state-of-the-art models estimating building energy consumption, however, rely on data that is costly or even impossible to collect at a large scale such as household income, dwelling characteristics and occupant consumption behavior patterns. This produces a gap between the demand for building energy consumption data (actual or estimated) and the supply. To fill this gap, we propose a method for estimating individual building energy consumption automatically from satellite or aerial (collectively, overhead) imagery data. It is impractical for human analysts to manually extract this information from large volumes of overhead imagery, and therefore we employ computer vision techniques to *automatically* extract the footprint of individual buildings, and then estimate their energy consumption. To our knowledge, this is the first study to directly use overhead imagery to estimate individual residential building energy consumption, potentially offering a powerful new tool for scalable and cost-effective data collection. However, this approach builds on a number of previous studies investigating methods for accurately estimating the energy consumption of buildings.

1.2. Previous work on estimating building-level energy consumption

Past work on building energy consumption estimation can broadly be divided up into two categories: *physical models* and *regression models*, with hybrid methods possible that combine both physical and statistical techniques [18]. Most of the existing building energy estimation methodologies require substantial amounts of data on each building to make effective estimates of energy consumption.

Physical models. One approach to analyzing the energy consumption of buildings is through the use of physical modeling and simulation techniques. Most physical models focus on analyzing thermal flows and are based on computational fluid dynamic (CFD), zonal and/or multizonal (nodal) models. Seminal works in this space include Clarke (2007) [19] and Underwood and Yik (2008) [20], which present wide-ranging discussions of these approaches. In the field of building energy consumption, thermal modelling is typically used together with a building energy model to simulate energy consumption, such as EnergyPlus [21]. *CFD models* describe thermal transfer in a building by producing a detailed description of thermal flows. This approach comes at a large computational cost and information requirement about the building and its materials. In the context of building energy consumption, these approaches have been applied for large rooms [22], predicting cooling and heating demand in an office [23], and improving overheating prediction through coupling CFD with a building energy simulation model, TRNSYS, for a typical Belgian two-story house [24]. *Zonal and multizonal models* divide rooms into several cells and solve physical equations for each cell. The focus of much of this work is on thermal comfort [25] and energy consumption estimation for buildings [26]. *Multizonal* approaches have been used to simulate building thermal behavior with each building component modeled as a separate node [27] and to model indoor air temperature in office buildings [28]. Olsen et al compare cooling systems using EnergyPlus in a nodal framework to evaluate energy savings potential for a newly built house in the UK [29]. Phan et al. develop a multi-zone modeling strategy for a data center housing 1,120 servers comparing energy performance between warm and cold climates [30]. Yang et al. combine a microclimate model ENVI-met with EnergyPlus to analyze the effects of different microclimatic factors on the energy balance of an individual urban building [31]. Scientists are also actively working on improving the quality of building energy consumption models. For instance, Dols et al couple EnergyPlus with CONTAM, a commonly used multizone building airflow simulation tool, aiming to capture the interdependencies between airflow and heat transfer [32].

Physical models, with the ability to attain unparalleled levels of modeling detail also require extensive information about each individual building, in some cases requiring a complete floorplan and inventory of building materials. These models may excel at making precise estimates but are not generally designed for large-scale implementation across many buildings in a region.

Regression models. Machine learning (or statistical learning) models that perform regression to estimate building-level energy consumption may be further subdivided into two subcategories: those that take a *top-down* approach (start with an estimate for the whole city/

region and disaggregate as needed) and those that take a *bottom-up approach* (start with an individual building estimate and aggregate up to a larger region) [33]. Top-down models often treat groups of buildings as a single energy-consuming collective, disregarding differences among individual buildings. Bottom-up models, on the other hand, focus on individual buildings and energy uses within buildings. Those energy estimates can then be aggregated upward. Li et al. review both approaches as well as variations within those approaches, advantages and disadvantages of each, and present an overview of the contemporary studies subscribing to each method [3].

The advantages of a top-down model for estimating building-level energy consumption are that this approach does not require the modeler to know buildings' actual energy consumption. In fact, this approach only requires limited input information, often with aggregated economic data, and may consider long-term economic effects. This approach does pose limitations, however. Modelers need long term historical data describing energy use, and the model's output relies upon those past relationships between economic factors and energy use. For deriving building-level energy consumption estimates, this typically amounts to using an average value or a weighted average of energy consumption for the region for each building without taking into account individual building attributes [34].

Bottom-up approaches may employ statistical models or physics-based models. Statistical models simulate energy use at the level of specific buildings, so variations among individual buildings are considered. However, this approach requires billing, weather, or survey data, plus a larger number of sampling subjects, and the output relies highly on historical energy consumption trends. Physics-based models simulate energy use at different temporal scales and consider variations in individual end use, but this approach requires detailed physical and technological measures as well as intensive computational effort.

Another branch of energy consumption estimation is energy forecasting – prediction of energy consumption in a time series rather than a cross-sectional setting. Examples of this approach include Li et al. who train a deep learning model to predict the energy consumption of a retail building at 30- and 60-minute intervals [35] and Dong et al. who used four years of monthly energy use data for commercial buildings in Singapore and use energy consumption history combined with weather attributes to predict future consumption [36]. These techniques rely on historical timeseries data, which is generally unavailable.

Most relevant to this work are bottom-up approaches that attempt to estimate individual building energy consumption. These techniques explore a wide variety of variables for predicting building energy consumption or identifying the most informative variables related to building energy consumption prediction. The variables considered in these studies range from weather, occupant lifestyle, dwelling size, and characteristics to socio-demographics [37–46]. However, these studies have one thing in common: they all rely upon numerous input variables, most of which are typically unavailable, thereby limiting the applicability of such models for large-scale building energy consumption prediction.

For instance, Sanquist et al. extract 5 lifestyle factors via factor analysis, which reflect social and behavioral patterns associated with air conditioning, laundry usage, personal computer usage, climate zone of residence, and TV use, from 17 variables in the U.S. Residential Energy Consumption Survey (RECS) to train a linear model capable of predicting building energy consumption with an R^2 of 0.4 [44]. Huebner et al. use Lasso regression and show that the following predictors are statistically significantly correlated with consumption: household size, length of heating season and all the physical building characteristics (e.g. floor area, fuel type, age, location) [37]. In addition, they train separate models for building energy consumption using combinations of groups of predictors. In particular, the physical building factors model performs best with an R^2 of 0.39, while the combined model that also includes socio-demographic and behavioral attributes such as number, age, gender, employment status and heating behavior of occupants has an R^2 of 0.44. Scott Kelly makes use of the English House Condition Survey and predicts via a linear model using 6 variables such as household size, income, and floor area to arrive at an R^2 of 0.31 [42].

Instead of predicting total energy consumption, some authors build models to predict energy use *intensity* or energy consumption *by end use*. For instance, Ma et al. use 216 variables and apply Elastic Net feature selection followed by Support Vector Regression to estimate use intensity for residential buildings in New York [38]. Whereas Min et al. utilize dozens of variables from RECS including energy price, household and housing unit characteristics to predict energy consumption by end use. Their log-linear models predicting heating, cooling, water heating and appliance energy consumption have R^2 values of 0.83, 0.70, 0.34, and 0.52 respectively [39].

Each of the building-level models discussed in this section rely on data that are generally unavailable, such as household size, income, or lifestyle information. Amasyali et al. [47] and Ahmad et al. [48] echo the insufficiency of data for effective energy consumption estimation and/or forecasting. This means these techniques would be restricted to spatial coverage of areas where such data are available. Fortunately, however, many of these studies also point to the physical size and characteristics that may be externally visible for a dwelling as being a strong predictor of energy consumption [37,39–41,49]. Our goal is to overcome the challenge of large-scale energy consumption estimation for residential buildings and to approach this we make use of the important observation that the size of a building (its footprint) is typically visible from overhead imagery, and so with the right tools to extract that information, we are able to generate an estimate of building energy consumption anywhere in the world with high resolution overhead imagery.

1.3. Previous work on the use of overhead remotely sensed imagery for energy-relevant studies

Machine learning and computer vision techniques, especially convolutional neural networks (CNN), have enabled many relevant remote sensing applications. For example, remotely sensed nighttime lights have been successfully used to evaluate electrification rates in Africa [50] and India [51] as well as to predict energy consumption in Japan, China and India (improving an R^2 from 0.66 to 0.83) [52]. The main limitation of nighttime lights dataset is its low resolution which limits its usefulness in analyzing building-level energy consumption. Related techniques have even been used to predict poverty in Africa using only remotely sensed data [53].

Daytime overhead imagery has also been shown to be effective for related analyses including solar photovoltaic array identification and capacity estimation. Building energy consumption may be impacted by the presence of solar photovoltaic arrays, so additional information on solar array location and size may help increase the accuracy of building energy estimation. Malof et al. demonstrate that solar photovoltaic panels can be detected extremely accurately with deep learning techniques [54], and Yu et al. demonstrate how this technique can be scaled up to the contiguous US [55].

Building detection and segmentation in overhead imagery is an active area of research. One example is Inria Aerial Image Labeling Benchmark (INRIA) released in December 2016–910 square kilometers of pixelwise labeled aerial imagery in 5 cities [56]. State-of-the-art building segmentation models generally use an encoder-decoder structure as described in Chen et al. [57], with skip connections to maintain the fine-grained object boundary details. Two popular variants of this framework are U-net [58], the network used for building segmentation in this paper, and D-LinkNet [59]. Some authors, on the other hand, use different common networks as an encoder and a decoder. For instance, Demir et al. [60] replace the feature extractor with larger pretrained deep neural networks like ResNet [61] arriving at a superior

performance on building segmentation benchmark datasets [62].

1.4. Focus on residential building energy consumption

Commercial building energy consumption tends to be better understood because of direct management for reducing energy consumption as compared with residential energy consumption which is highly variable due to variations in building sizes and geometries, occupant behavior, and limited data [33]. For many residences, the set of potential appliances or devices that are present may vary based on the culture [63] as well as the number and affluence of the occupants [21]. In this work, we focus on residential buildings due to two factors: (1) from a practical perspective, data on individual residential building energy consumption (including precise location and energy consumption) are publicly available for multiple cities, while very little information that is not de-identified is publicly available for the commercial sector; and (2) residential and commercial building energy consumption collectively accounts for 40% of total U.S. energy consumption, split about equally between the two sectors, so there is a significant impact for each. Moreover, residential buildings account for a large proportion of global energy consumption in both low and high income countries making this study topical outside the U.S. as well [64]. While there has been some work on commercial building energy consumption estimation [65–67], these approaches typically also require detailed information about a building. We leave the investigation of similar satellite imagery-based energy consumption estimation techniques that do not require such detailed data for each building to future work.

1.5. Contributions of this work

In this paper, we introduce and evaluate the performance of a novel approach for estimating individual and regional residential building energy consumption directly from overhead imagery, without the need for detailed information on a building or any historical or demographic data related to the building. This concept, originally suggested in [49], relies on deep convolutional neural networks to automatically estimate building footprints, classify buildings by type (residential or commercial) and extract relevant features (e.g. building footprint area, perimeter, population density) in color overhead imagery. To these features extracted for each individual building, we apply a regression model to estimate that building's energy consumption. We demonstrate that our approach can efficiently scale to analyses over large geographic areas, while still providing consumption estimates at a high geospatial resolution (e.g., individual buildings or neighborhoods), and that this model can be applied in multiple locations (in this work Gainesville, Florida, and San Diego, California).

The paper is structured as follows: we summarize the datasets used in this study (Section 2) before describing the methodology (Section 3) for estimating building energy consumption. The methodology describes how we detect buildings in overhead imagery and generation of building footprints (3.1); classify those buildings as residential or commercial buildings (3.2); extract features based on those buildings including quantifies such as building area and perimeter (3.3); and predict the energy consumed by the building using regression (3.4). We also discuss the particular scoring metrics used to compare across models (3.5) and the baseline estimator we adopt as a point of comparison for performance (3.6) and how we split our data into training and testing sets (3.7). We lay out our experimental design in Section 4 to test the performance of each component of our analytical pipeline including building segmentation (4.1); building type classification (4.2) and energy consumption estimation (4.3 and 4.4). We also perform a sensitivity analysis to the quantity of training data needed (4.5). We present the results for each of these experiments in Section 5 (with each subsection in Section 5 corresponding to the experiment with the same number in Section 4) before concluding. Fig. 1 describes the complete pipeline for the methodology in this work and directs the reader to the relevant section containing experimental results for each pipeline component.

2. Datasets

Data on building-level energy consumption are sparse and we need data where we have both building-level energy consumption *and* high-resolution overhead imagery. Note that in practice, building-level energy consumption information is not necessary to use our proposed method, but it is necessary to evaluate the ability of our approach to estimate the correct values (i.e., measure its accuracy). The two suitable areas we found are Gainesville, Florida, and San Diego, California. In both of these locations, high-resolution overhead imagery and energy consumption data are available, although the resolution of the energy data varies by region. The available data for each region are summarized in Table 1.

Overhead imagery data. High-resolution, 0.3 m aerial imagery is publicly available from the United States Geological Survey (USGS) for both Gainesville and San Diego from the high resolution orthoimagery collection [69]. For building identification, 0.5 m or finer resolution data are typically required for recent state-of-the-art building segmentation algorithms, so 0.3 m resolution is adequate. The USGS imagery data were collected in December 2013 for the Gainesville data and October 2014 for the San Diego data.

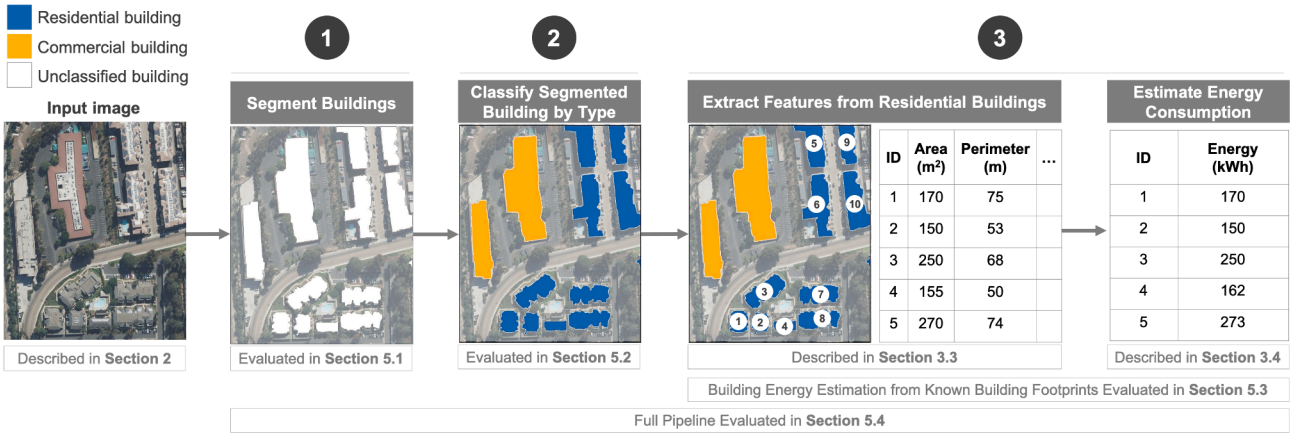

Fig. 1. Flowchart of the full pipeline of our methodology. Start with only an input satellite image, then (1) detect and segment each building's footprint, (2) classify the building by type (residential/commercial), (3) extract relevant features based on the data, and finally use those features to estimate building energy consumption. Each item above indicates where in this manuscript additional information about the process can be found.

Table 1
Description of data availability. For Gainesville, Florida, we have both energy consumption and building-level footprints for residential buildings. In San Diego, CA, we have ZIP Code level energy consumption and anonymized individual building energy consumption data. The experiments in this work are designed to maximize the use of these data sources and to do so with as much logical symmetry in approach as is possible given the constraints of the data.

Location	Number of individual buildings with known energy consumption	Individual building area (in square meters) known?	Includes corresponding building footprints (polygons)?	Includes ZIP Code-level energy consumption?	Additional associated data
Gainesville, FL	26,991	Yes	Yes	No	A total of 33,432 building footprints from which to extract ground truth derived predictors; hand-labeled building footprints for 1,098 buildings[5] (for fine-tuning a building segmentation algorithm pretrained on INRIA building data [56] and for testing the algorithm)
San Diego, CA	10,718	Yes	No	Yes (90 ZIP Codes)	Hand-labeled building footprints for 7,626 buildings[1] (for fine-tuning building segmentation algorithm used in Gainesville and for testing the algorithm); zoning information for San Diego City

[1]A subset of images was manually annotated using pyimannotate [68]. The annotator would mark each building's perimeter manually.

Energy consumption and building footprint data. In Gainesville, Florida, we downloaded a detailed energy consumption dataset provided by Gainesville Green containing monthly electricity, water, and natural gas consumption as well as the corresponding geographic location for over 30,000 households in 2016 [10]. The data includes other years as well that were not included in this study. In Alachua County, in which Gainesville resides, the Property Appraiser made available a geospatial dataset of most building footprints in the county. Merging and matching the building footprints data with each building's corresponding energy consumption, resulted in a dataset of 26,991 households in Gainesville with the corresponding monthly energy consumption [70]. For each building, we aggregated monthly energy consumption up to annual consumption and removed any buildings with zero consumption (these are likely not in active use).

In San Diego County, the available energy data were not fully-identified in the same way that they were in Gainesville, so we gathered data from three sources. The first dataset includes quarterly ZIP Code level energy consumption data publicly available from San Diego Gas & Electric [71]. This quarterly dataset is divided up by customer class that provides consumption by a number of building types including residential and commercial buildings. We extracted the residential data and summed the quarterly data to calculate the annual ZIP Code level energy consumption. The second dataset is a sample of individual residential building-level energy consumption data from 10,718 location-anonymized buildings in San Diego City from the Building Performance Database [72]. While this dataset does not contain information on the particular location of the building (so we cannot match it directly with imagery data) it does contain information on the building square footage (which while not the same as the building footprint area, would be the same for 1-story buildings) and their ZIP Code location. We can use this information to train a regression model on and evaluate individual building-level energy consumption prediction performance, and its corresponding ZIP Code. Lastly, we have zoning data in San Diego City [73] to provide information as to whether a building in San Diego is residential or commercial.

3. Methodology

To predict residential building energy consumption from overhead imagery, we divide this process into three steps, illustrated in Fig. 1: (1) building detection and segmentation, (2) building type classification, and (3) energy consumption prediction, where we estimate the energy consumed by that building based on features of the building. Step (1) detects buildings in overhead imagery and produces polygonal representations of each detected building, a process known as building segmentation [74]. Residential buildings being the focus of the present paper, step (2) classifies the building as either being a residential or commercial building. Step (3) extracts informative features from the building footprints of residential buildings (e.g. area, perimeter) and their surrounding context (e.g., population density) to train our energy consumption prediction model and generate estimates of the energy consumed by each residential building.

3.1. Building segmentation

The first step in this process is to analyze the overhead imagery and identify where each building is located and generate a polygon denoting its outline. This is done by estimating the specific pixels in an image that compose each building in a process that is known as image segmentation [75]. The result of segmentation in this case is a binary image where each pixel is labeled with a value of one if it resides on a building, and a value of zero otherwise. This process can be done via manual labeling, however, given the size of overhead imagery, this approach is costly and impractical for regular application. Recently, it has been shown that convolutional neural networks (CNNs), which are a special type of machine learning model, can be *trained* to automatically generate high-accuracy image segmentation of various objects in overhead imagery, including buildings. Once such a model is trained, it is capable of rapidly scanning vast quantities of overhead imagery, making it scalable and cost-effective. This approach is ideal for our application, which requires regularly segmenting large volumes of overhead imagery to obtain up-to-date energy consumption estimates.

Training a CNN requires providing it with a set of image pairs, where each pair comprises one overhead image and the desired segmentation map for that image. CNNs are composed of a large number of model parameters that control its segmentation process. During training, the CNN will repeatedly be provided with training images, and then attempt to predict their segmentation maps. The CNN's parameters will be automatically and gradually adjusted throughout this process to maximize the agreement between its predictions on the training imagery and their corresponding segmentation map. CNNs typically require a large set of training imagery, which themselves usually must be manually annotated [76]. However, this process must only be done once, after which the CNN can label much larger quantities of new imagery. Furthermore, there are now publicly available datasets to train segmentation models, including those for building segmentation. In this work we initially train our segmentation models on the INRIA dataset, which includes high-resolution orthorectified overhead color imagery covering 405 km^2 of area across 5 major cities in the US and Europe with full segmentation labels for training. After we complete training on INRIA we subsequently train the same CNN further on a smaller set of imagery from Gainesville and San Diego, where we ultimately apply the model for segmenting new imagery. This process of training a second time on task-specific imagery is known as "fine-tuning", and it has been shown to enhance the performance of CNNs for recognition tasks [77]. More details of the training setup can be found in [56].

Segmentation CNNs are available in many different architectures, each with somewhat varying characteristics. In this work we employ the U-net architecture [58], which has demonstrated success especially for building segmentation [58,78]. In particular, we use a modified U-net architecture that recently achieved the best performance for building segmentation on the INRIA dataset [56]. We trained our U-net as described above and then subsequently used it to segment all the buildings in our target imagery over Gainesville and San Diego. The segmentation output of our U-net is a binary image, however, we ultimately need to identify the polygons corresponding to individual buildings. To achieve this we extracted contiguous groups of building pixels – termed connected components – and then applied the Douglas–Peucker algorithm [79] to approximate a polygon that matches well with the shape/size of the connected component.

3.2. Building type classification

Building segmentation produces polygons that may represent many types of buildings including both residential and commercial. Therefore, we need to be able to identify those buildings that are residential before we can assign energy consumption estimates to those buildings. To do so, we trained a building type classifier that takes as input an overhead image of a building (cropped to 224 × 224 pixels around the center of the building) and classifies each building or rather the 224 × 224 pixel neighborhood around each building as either residential or commercial. The building type classifier is based on ResNet-152 architecture [61] (a well-known architecture designed for image classification) and assigns a residential or commercial class to each building.

3.3. Feature extraction

For each residential building identified, we know the location in the overhead image via the polygon (outline) of that building produced from the segmentation step. From this, we need to extract meaningful features from the imagery and polygon that inform a prediction of the amount of energy consumed within the building. Inherent in the building polygon and corresponding imagery there may be valuable information. Area (a measure of building size) and perimeter (relative to area is a measure of building complexity) are directly computable for each object giving us insight into the physical characteristics of a building. Area is an intuitive predictor for homes with significant heating and cooling energy consumption [80].

Other features may be worth considering, including information on rooftop material or swimming pools, population density (number of neighboring buildings within certain radius of the building), and other neighborhood characteristics. As mentioned before, CNNs are particularly adept at automatically extracting meaningful features, so we can also allow the CNN to automatically attempt to determine those properties which are of greatest value for estimating energy consumption—some of these may have semantic meaning (e.g. roof color, presence of two cars in a driveway)—while others may not have such interpretability. We do this using ResNet-152 model [61]. We apply the ResNet model pre-trained on ImageNet [81] only for the purpose of extracting features, which are the network values fed into the final fully connected layer of the ResNet CNN. We then apply principal components analysis to these 2,048 features and use the top 10 principal components of those ResNet-extracted features as additional features about the building and its surrounding neighborhood [82]. While these additional features do contain information about the buildings, we show in the analyses below that simple building area proves to be, overall, the most informative feature for this problem.

3.4. Energy consumption prediction

Using the features extracted from the imagery data, area in particular, we explored a number of regression models for energy consumption prediction that have been prominent in the literature [10,11]: linear multiple regression, gradient boosted decision trees, artificial neural networks, and random forests [49]. Using the model which predicted individual building energy consumption prediction best, random forests, we investigated the most informative features for this prediction by comparing performance across different combinations of features.

This prediction approach estimates individual building energy consumption. The use of energy in individual buildings may be extremely variable due to factors that are invisible in overhead imagery including electronic devices that are not related to building size. However, small amounts of spatial aggregation may significantly reduce this variability while still retaining a high level of geographic spatial resolution. We therefore investigate the impact of spatial aggregation on building energy estimation accuracy with an aggregation range of one square kilometer or less, much smaller than most ZIP Codes and neighborhoods. Past work in automatic solar photovoltaic identification demonstrated that geospatial resolution may be traded for estimation accuracy [83], and we also test the efficacy of that approach for building energy estimation.

3.5. Scoring metrics

To evaluate the performance of each regression model we use two well-known scoring metrics, which are two different ways to quantify the error between the energy consumption predicted by the models, and the true energy consumption values. The first is the coefficient of determination, R^2, which is the proportion of variance of the true energy consumption values that are explained by our model's predictions; we want this measure to be as close to 1 as possible (representing all variance being explained by the model). We use the common formulation for R^2, defined as follows

$$R^2(\boldsymbol{y}, \widehat{\boldsymbol{y}}) = 1 - \frac{\sum_{i=1}^{N}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{1}$$

Here y_i is the true value, $\widehat{y_i}$ is the predicted value of energy consumption for the i^{th} building and $\bar{y}$ is the sample average over true values.

The second is the root mean square error (RMSE) between the predicted and true energy consumption values, given by

$$RMSE(\boldsymbol{y}, \widehat{\boldsymbol{y}}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2} \tag{2}$$

We seek models that minimize the RMSE, with zero being optimal. For aggregation regions we simply compute R^2 and RMSE on the aggregate truth and predictions, where y_i would then be the true value of energy consumption for the region and $\widehat{y_i}$ is the predicted energy consumption value for that region.

3.6. Baseline building energy consumption estimator

There are two possible approaches for comparing the efficacy of our method. The first is to use existing building energy consumption estimation models, but as previously discussed those models often rely on building-specific data on the contents of the building or owner demographics that are rarely available at the individual building level. The second approach is to use regional energy consumption averages. In the absence of high-resolution geospatial energy consumption data, regional or national estimate of average building energy consumption is often available to policymakers and researchers. Therefore, for this work we adopt a regional average of building energy consumption as our baseline estimator for building energy consumption.

The next choice is between using a national or regional average residential building consumption estimate. The rationale for a regional

rather than national average is due to the regional variation on building energy consumption. Data on this variation can be challenging to find. Although average annual energy consumption for some cities and counties is published [84,85], there are no comprehensive databases that collect city-level average consumption across the entire U.S. However, there are some data sources one could use for a geographic comparison. The Residential Energy Consumption Survey (RECS) is an example data source for regional estimates of average residential consumption, with regional resolution at the census level. This database demonstrates significant regional differences. For instance, while the national average building energy consumption from RECS 2015 is 11,028 kWh, in New England the average is 7,633 kWh, while in parts of the southern U.S. the average is 14,807 kWh.

For the purposes of our experiments we use a different baseline average estimate for each of the two regions in this study since this approach would generally be more accurate than using a single national average. In Gainesville, we compute the average for the training set (11,723 kWh) and use it as the baseline prediction for each building in the test set. In San Diego, we use the county average (5,871 kWh) [86]. Some of our experiments also explore the impact of spatial aggregation on estimating energy consumption. In the limit as the aggregation area approaches the extent of the region over which the original average was based on, the average becomes a perfect estimator. Since these regional estimates are not available for all cities and certainly not available at the neighborhood or individual building-level, we aim to show that our model outperforms this baseline estimator in most cases.

3.7. Training and testing data

To prepare our Gainesville data for machine learning experimentation, we divided our previously described energy consumption data into a training and test dataset, where the training data accounted for 75% of the samples. Since we conduct experiments for both the individual buildings and buildings aggregated over small regions, we took care to prevent information from our training data leaking into our test data. Therefore, we split the training and test data randomly, ensuring that held-out test data are never seen during training. We also performed the random sampling in a way such that we grouped neighborhoods together in gridded regions, sampling those regions into either the training or test set. The goal here was to reduce the likelihood that two neighboring houses would appear one in the training set and one in the test set. We also use group 5-fold cross validation on the training data for all of our experiments to select any model hyperparameters. Similar to the Gainesville dataset, for San Diego we randomly split the building-level data with 75% of the samples in the training dataset and 25% in the test dataset based on each buildings' ZIP Code location. Anonymized building-level energy consumption data in San Diego comes from 16 ZIP Codes. To avoid information leakage in San Diego we do not test on the 12 ZIP Codes from which we had training data.

4. Experimental design

Since the building energy estimation approach described here is composed of a number of subcomponents: (1) building segmentation, (2) building type classification, and (3) energy consumption estimation, we first create three brief experiments to individually measure the efficacy of each of these steps to evaluate potential sources of error that may be introduced when combining these modules.

4.1. Building segmentation evaluation

To evaluate how well our approach could segment individual buildings, we used 798 hand-labeled building footprints in Gainesville and 7,396 from San Diego that we collected to serve as training data[1]. We trained our building segmentation algorithm and tested on a held-out dataset of 300 buildings in Gainesville and 230 buildings in San Diego, resulting in the performance demonstrated in Fig. 2. These precision-recall (PR) curves demonstrate the performance evaluated on a held-out set of images for both Gainesville and San Diego, quantifying how well individual buildings can be segmented in two diverse regions.

For the PR curves, recall is the fraction of buildings in the ground truth that were correctly detected. Precision is the fraction of objects in the image we estimated to be buildings that actually were buildings.

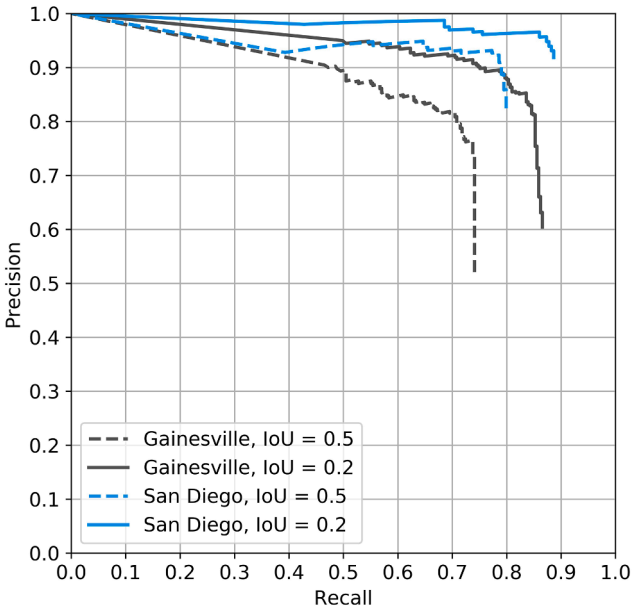

Fig. 2. Precision-Recall performance curves for building segmentation from overhead imagery. Comparison of object-wise Precision-Recall (PR) curves for Gainesville and San Diego. Each object is deemed a detection, i.e. a true positive, if the Intersection over Union (IoU) of the segmented object compared to the ground truth exceeds a threshold value (IoU thresholds of 0.2 and 0.5 are used to score segmentation quality in both locations). An IoU of 0.5 is more restrictive on the quality of the segmentation that is required for a detection to be considered a true positive rather than a false negative; a threshold of 0.2 enables more buildings to be declared as "detected," albeit with less precise segmentation.

[1] Training data are required for building segmentation algorithms, and even with a pretrained algorithm, when approaching new geographies examples from that new geography improve performance. In this case, although we have building footprints in Gainesville, the geographic coordinates of objects in overhead imagery may not align precisely with objects of significant height (such as some buildings) since the camera angle may shift where a building rooftop appears in the image. This can be seen clearly when viewing skyscrapers from above. To overcome this issue, we gathered training data based on the specific imagery we were using, requiring manual annotation of those buildings.

These curves are evaluated object-wise; that is each building detected could be either a true positive or false positive. A building was scored as a true positive if the detected building and ground truth building overlap sufficiently. Overlap is defined by intersection-over-union (IoU) which for two polygons represents the area of polygon intersection divided by the area of the union of the polygons. An IoU of 1 represents complete overlap, an IoU of 0 represents no overlap. Once we know how to identify true positives, we can then vary the sensitivity of the detector to produce each PR curve. Recall is the fraction of actual buildings that were labeled as buildings. Precision is the fraction of those objects that we labeled as buildings which were actually buildings (so low precision would mean there were many false positives). Precision and recall of 1 would represent perfect performance.

4.2. Building type classification evaluation

To evaluate how well our approach classifies individual buildings as residential or commercial, we evaluate the classification performance in San Diego City. In Gainesville we only have residential building data so that location would not be an ideal location for testing this part of the pipeline. In San Diego, however, we do need to make this distinction. Another challenge is in having a sufficiently large dataset for evaluation since there are so many more residential buildings than commercial. To do this, we combine building footprint estimates produced by our algorithm and the zoning information for San Diego City. This combination allowed us to assign the label of residential or commercial to each detected building, creating the dataset described in Table 2 that has 189,448 buildings in the training dataset and 33,432 buildings in the test dataset with thousands of examples of both residential and commercial buildings. The massive class imbalance between residential buildings (32,047) versus commercial buildings (1,385) makes this problem more difficult for identifying commercial buildings, but the size of the dataset enables us to better evaluate our classifier's performance which we present in a confusion matrix form in section 5.

Due to the asymmetry of the available data between Gainesville and San Diego, the experiments we describe below use Gainesville data to evaluate performance at the individual building-level *and* at different levels of aggregation, while the San Diego experiments are only evaluated at the aggregated ZIP Code level. However, the same underlying process, that of Fig. 1, is used in both cases, and since the cities are separated by about 3,700 km, the results present two very different test cases that provide evidence towards the ability to apply this algorithm to geographically diverse areas.

4.3. Building energy consumption estimation process evaluation from ground truth building footprints

Before using the segmented and classified buildings, it is important to evaluate the performance of the energy estimation regression model using actual ground truth building footprint data. We use the ground truth building polygons we have in Gainesville and extract the building footprint for each building in our dataset. We then use building characteristic features extracted directly from the ground truth building footprints (e.g. area), so there is no error from building segmentation or type classification. Using this, we train an algorithm to predict building energy consumption using random forests regression. We apply the trained model to our test dataset and evaluate performance.

We evaluate performance at both the individual building level as well as at five levels of aggregation using grid cells of progressively increasing size: 50 × 50 m, 100 × 100 m, 200 × 200 m, 400 × 400 m, 1000 × 1000 m. We compare performance against the baseline of using the average building energy consumption as the estimate of each building (see Section 3.6).

Table 2
Overview of San Diego City building type classification datasets. Training and test datasets by the number of buildings contained of each type.

	Total	Residential	Commercial
Training Data	189,448	182,564	6,884
Testing Data	33,432	32,047	1,385

4.4. End-to-end process evaluation: Estimating building energy consumption from overhead imagery

After investigating the performance of each component of this process individually, in this experiment we evaluate the full model end-to-end, summarized in Fig. 1, using overhead imagery directly to identify, segment, and classify all buildings and then use building characteristics extracted from the imagery (e.g. building area) to estimate building energy consumption.

We apply this same process in both Gainesville and San Diego, although due to the nature of the data available for each, we evaluate performance differently for each location. In Gainesville, we repeat the process of Section 4.3, except with the full pipeline[2] and evaluate performance at both the individual building level as well as at five levels of spatial aggregation. In San Diego, however, since the individual buildings in the training data are location-anonymized and are only usable at the individual or ZIP Code levels of aggregation. We include San Diego to demonstrate performance in another metropolitan area and in a larger area (as the San Diego data covers both the city and the county) and is more relevant for large-scale regional aggregation to the ZIP Code level. In both regions, we compare performance against the baseline of using the average building energy consumption as the estimate of each building (see Section 3.6).

4.5. Sensitivity analysis to identify the amount of training data required

Since the energy consumption estimation methods proposed here rely on training data which includes examples of buildings with their location (latitude, longitude) and corresponding energy consumption, we wanted to evaluate the quantity of data necessary to achieve the performance one might target for each model and the consistency of that performance. To that end, we ran 100 trials of the end-to-end pipeline experiment (Section 4.4), estimating building energy consumption, varying the amount of training data used in each (sampling with replacement), but keeping the held-out test set fixed. We varied the size of the training dataset logarithmically from about 10 to 10,000 buildings in San Diego, about 20 to 15,000 buildings in Gainesville (since more buildings with known energy consumption were available in Gainesville).

5. Results

In this section, we describe the results from each of the experiments from the corresponding subsection in experimental design covering each of the subcomponents: (1) building segmentation, (2) building type classification, (3) energy consumption estimation, (4) the end-to-end experimental design, and (5) a sensitivity analysis of our results to the size of the training dataset.

5.1. Building segmentation results: How accurately are buildings detected and segmented in overhead imagery?

The results from segmenting buildings from overhead imagery are shown in Fig. 2 through precision-recall (PR) performance curves (explained in Section 4.1). We evaluated objects detected by setting an IoU threshold of both 0.5 (stricter in terms of segmentation quality) and

[2] Gainesville data only contained residential buildings, so building type classification was unnecessary.

0.2 (less strict) which requires excellent correspondence between the estimated and predicted polygons. In general, Gainesville buildings were more difficult to identify due to significantly more tree cover over buildings than in San Diego.

For the 0.5 IoU case In Gainesville we were able to correctly identify 74% of buildings with 24% of identified buildings being false positives. For San Diego the performance was higher (in part because we used the Gainesville model as a pre-trained model for San Diego); the segmentation algorithm correctly identified 78% of buildings with 8% of identified buildings being false positives. The average IoU was 0.81.

For the less stringent IoU threshold of 0.2, which would include a sufficient number of detected buildings while still preventing the inclusion of many spurious groups pixels that happened to be labeled buildings. In Gainesville we were able to correctly identify 84% of buildings with 15% of identified buildings being false positives. For San Diego we correctly identified 88% of buildings with 8% of identified buildings being false positives. The average IoU was 0.76.

5.2. Building classification results: How accurately are buildings categorized as residential or commercial?

The results from investigating the building type classification algorithm performance (described in Section 4.2) are shown in the confusion matrix in Table 3. This demonstrates that the classifier correctly identifies residential buildings 99% of the time and commercial buildings 74% of the time.

5.3. Building energy consumption results: Estimating building energy consumption from ground truth building footprints

The results from estimating building energy consumption (the experiment described in Section 4.3) directly from Gainesville ground truth building footprints are shown in Table 4 in two parts, labeled **A** and **B**. The **A**-columns represent the baseline results: regional average building energy consumption multiplied by the total number of actual buildings (Section 3.6). The underlying data used for performance evaluation is the entire test set and the estimator is the average of the training set. Column **B** shows the results from repeating the analysis for column **A**, except using the random forest regression algorithm trained on the ground truth area of each building in the dataset. Comparing the case without aggregation, the RMSE improves by 15.3% over baseline and increases with aggregation. For the 1000 × 1000 m case, the RMSE improved 37.1% over the baseline.

5.4. End-to-end process results: Estimating building energy consumption from overhead imagery

This section describes the results from applying the full energy consumption estimation pipeline (Fig. 1), using only overhead imagery, initially described in Section 4.4. The results are presented by region: Gainesville in Table 5 and Table 6 (see supplemental information for additional results), and San Diego in Table 7. As a reminder, for Gainesville, we have individual building *and* aggregate building energy consumption data with geolocation while for San Diego, we have only aggregate ZIP Code level energy consumption data, so the results are presented to reflect that difference.

As we evaluate performance for the whole pipeline, we are faced with an additional complication: the fact that not every building is going to be detected by our segmentation algorithm in the overhead imagery. For maximum demonstration of performance, we therefore show results for two conditions: (1) where we evaluate the energy consumption estimates for all buildings, acknowledging that some of those buildings will be missed by the segmentation algorithm and setting the estimate for those buildings to 0 kWh (Table 5) and (2) where we evaluate the performance only on buildings that were detected by the segmentation algorithm (Table 6). The latter case answers the question: "if we can identify that there is a building, how accurately can we estimate its energy consumption?" Overhead imagery may be occluded (such as by tree cover) making segmentation and detection of buildings challenging. Gainesville is an excellent example of this phenomena. Nevertheless, building segmentation techniques are being rapidly developed, so each of these approaches offer insights into algorithm performance and the potential for improvements in the near future.[3]

Setting buildings not detected to 0 kWh. The Gainesville results evaluated on all buildings while setting energy consumption estimates of buildings missed by the segmentation algorithm to 0 kWh are shown in Table 5. Column **A** is the baseline model and columns **B** and **C** represent applying the full pipeline (Fig. 1) trained on either the ground truth area of the buildings (column **B**) or on the estimated area from the segmentation algorithm (column **C**). By using the segmentation area rather than the ground truth area, the energy estimation regression model is able to learn to compensate for some of the bias inherent within the building detection process connecting the imperfect building footprints estimated by the segmentation algorithm the corresponding energy consumption

Table 3
Confusion matrix for San Diego building type classification. The central four cells contain the percent of buildings of one type (rows) predicted as being either residential or commercial (columns). These central bolded cells are percentages of the total number of residential or commercial buildings. The number of samples in the right column is the total number of samples of a particular category (so there are 32,047 residential buildings in the dataset and 1,385 commercial buildings).

		Prediction		
		Residential	**Commercial**	Number of samples
Truth	**Residential**	**99%**	**1%**	*32,047*
	Commercial	**26%**	**74%**	*1,385*
	Total Samples			*33,432*

Table 4
Results for building energy estimation from ground truth building footprints in Gainesville, Florida. Comparison of building energy estimation regression by predictors used and level of spatial aggregation for all buildings (those detected by the CNN and those missed, for 26,991 buildings in total) in Gainesville, Florida. All predictors are estimated using ground truth building footprints. The random forest model outperforms the average baseline, spatial aggregation significantly increases performance of all models. Area is a very strong predictor. The RMSE values are in units of kilowatt-hours (kWh). The baseline column is the point of comparison for performance improvement measurements across the table.

Regression Model	**A** Average (**baseline**)		**B** Random Forest		
Predictors	Training set average consumption		Ground Truth Area		
Level of aggregation	R^2	RMSE (kWh)	R^2	RMSE (kWh)	**Improvement over RMSE baseline (%)**
None (Individual)	0.00	6541	0.28	5537	**15.3**
50 × 50 m	0.22	6854	0.48	5611	**18.1**
100 × 100 m	0.55	7399	0.74	5495	**25.7**
200 × 200 m	0.74	7252	0.88	5100	**29.7**
400 × 400 m	0.81	6281	0.93	4044	**35.6**
1000 × 1000 m	0.87	6590	0.96	4143	**37.1**

[3] We also explored energy consumption prediction performance enhancements from imputing the values of missing buildings, if we assume we know how many buildings are in a given region, more information can be found in the supplemental information. This demonstrated that additional performance improvements if we know how many buildings are in a given region, even though we do not know their particular characteristics.

Table 5
Results for the end-to-end building energy estimation process for Gainesville with missed buildings set to 0 kWh estimates. Comparison of building energy estimation regression by predictors used and the level of spatial aggregation for all 26,991 buildings in Gainesville, Florida with predicted consumption of buildings not detected by the segmentation algorithm set to 0 kWh. Column A is the baseline performance of the average for all buildings but with buildings missed by the segmentation algorithm set to 0 kWh. The RMSE values are in units of kilowatt-hours (kWh). The baseline column is the point of comparison for performance improvement measurements across the table. Columns B and C depict performance of random forest algorithm trained on ground truth (column B), CNN-estimated area (column C), and CNN-estimated area with additional features included (column D).

Regression model	**A** Average (Baseline with consumption of buildings missed by segmentation set to 0 kWh)		**B** Random forest			**C** Random forest			**D** Random forest	
Predictors	Training set average consumption		Area training predictors: ground truth area, test predictors: estimated area			Area training AND test predictors: estimated area			CNN estimated area, perimeter, local density estimates, top 10 principal components of the ResNet-152 features	
Level of aggregation	R^2	RMSE (kWh)	R^2	RMSE (kWh)	**Improvement over RMSE baseline (%)**	R^2	RMSE (kWh)	**Improvement over RMSE baseline (%)**	**Improvement over RMSE baseline (%)**	
None (Individual)	0.00	8199	0.00	7965	**2.9**	0.00	7900	**3.6**	**7.2**	
50 × 50 m	0.00	8349	0.00	7931	**5.0**	0.00	7895	**5.4**	**9.4**	
100 × 100 m	0.33	8428	0.27	7530	**10.7**	0.35	7691	**8.7**	**16.8**	
200 × 200 m	0.64	8275	0.60	7275	**12.1**	0.66	7268	**12.2**	**23.4**	
400 × 400 m	0.72	7140	0.68	6237	**12.6**	0.75	6005	**15.9**	**30.7**	
1000 × 1000 m	0.77	7343	0.76	6215	**15.4**	0.81	6021	**18.0**	**35.3**	

Table 6
Results for the end-to-end building energy estimation process for Gainesville for CNN-detected buildings. Comparison of building energy estimation regression by predictors used and by the level of spatial aggregation for the 20,122 buildings detected by the segmentation algorithm in Gainesville, Florida. Predictors used are ground truth area (column B) or are estimated from CNN-derived building annotations (columns C and D). The Random Forest prediction outperforms the average baseline in all scenarios. Adding additional predictors to building area (column D) is strictly better than using estimated area alone (column C), while it also outperforms energy predictions using ground truth area (column B) for larger levels of aggregation. The RMSE values are in units of kilowatt-hours (kWh). The baseline column is the point of comparison for performance improvement measurements across the table. See Table 9 in the Supplemental Information for additional performance metrics for column B.

Regression model	**A** Average (**baseline**)		**B** Random forest			**C** Random forest			**D** Random forest		
Predictors	Training set average consumption		Ground truth area			CNN estimated area			CNN estimated area, perimeter, local density estimates, top 10 principal components of the ResNet-152 features		
Level of aggregation	R^2	RMSE (kWh)	**Improvement over RMSE baseline (%)**	R^2	RMSE (kWh)	**Improvement over RMSE baseline (%)**	R^2	RMSE (kWh)	**Improvement over RMSE baseline (%)**		
None (Individual)	0.00	6851	**17.9**	0.14	6352	**7.3**	0.28	5831	**14.9**		
50 × 50 m	0.27	7147	**20.3**	0.35	6472	**9.4**	0.48	5912	**17.3**		
100 × 100 m	0.61	7604	**26.5**	0.68	6716	**11.7**	0.76	5859	**22.9**		
200 × 200 m	0.79	7257	**30.5**	0.86	6101	**15.9**	0.91	5009	**31.0**		
400 × 400 m	0.85	6264	**35.1**	0.92	5040	**19.5**	0.95	3853	**38.5**		
1000 × 1000 m	0.87	6546	**36.1**	0.94	5177	**20.9**	0.97	3784	**42.2**		

of the building. We see that both models as a whole improve upon the average baseline ranging from around 2–3.6% for individual buildings and 15–18% for the 1 km^2 aggregation, with the larger gains occurring when the model is trained on the building area estimated from the segmentation results.

As described in Section 3.3, we can also extract additional features from overhead imagery relevant to each building. In addition to the area estimated through the segmentation process in Section 3.1, we also extract features including perimeter, local building density estimates (a measure of how many other buildings are nearby), and the top 10 principal components of the ResNet-152 features. Adding these features to the random forest with these additional features (column **D** of Table 5[4]) nearly doubles the performance gains against the baseline: a 7.2% RMSE improvement for individual building and a 35.3% improvement for 1000 × 1000 m aggregation, compared to the 3.6% and 18% improvement that random forest trained on area alone achieves, respectively.

In Section 5.3, where we only considered ground truth building footprint area, that represented the best-case scenario for area estimates. This model (Table 5.**D**) outperforms the baseline even in that ideal case. Using the additional features described here, the RMSE of building energy consumption for regions at least 100 × 100 m in size (7,009 kWh, full table available in Supplemental Information) is lower than the corresponding baseline estimate from Table 4 (7,399 kWh) which uses the actual building energy consumption average as well as ground truth footprints. Even with the imperfect building detections, with the right predictors this energy consumption estimation technique can improve over a baseline that is aware of the average building energy consumption and the location of all buildings.

Limiting performance evaluation to detected buildings. Our next set of results address the question: "How well do we estimate building

[4] See Supplemental Information Table 10 for additional performance metrics

Table 7
Results for end-to-end building energy consumption estimation for San Diego. ZIP Code-level prediction performance using CNN-estimated annotations classified as residential, then applying Random Forest regression and comparing against the averages for either San Diego county (baseline 1) or San Diego City (baseline 2). The results for San Diego City (18 testing ZIP Codes, column A) and San Diego County (78 testing ZIP Codes, column B) are presented separately as the former is a higher density region and the random forest energy consumption regression model was trained on an anonymized sample from San Diego City. ZIP Code sizes range from 0.16 sq. km to over 1,700 square km and are on average about 100 sq. km. All 12 ZIP Codes where anonymized training data comes from are excluded from testing.

Location	**A** San Diego City					**B** San Diego County		
Number of ZIP Codes tested (trained)	18 (12)					78 (12)		
Prediction method (random forest or average)	Average (county, baseline 1)	Average (city, baseline 2)	Random Forest	**Improvement over Baseline 1 (%)**	**Improvement over Baseline 2 (%)**	Average (county, baseline)	Random Forest	**Improvement (%)**
R^2	0.36	0.55	0.77	114	40	0.72	0.75	4.2
Correlation	0.82	0.82	0.93	13.4	13.4	0.91	0.93	2.2
RMSE (kWh)	5671	5410	3737	**34.1**	**30.9**	10,404	9742	**6.4**
Average number of buildings per ZIP Code	9,333					9,171		
Total number of buildings	167,997					715,353		

energy consumption assuming we identify the building in overhead imagery?" For this case, we only consider Gainesville results for buildings identified by the segmentation algorithm, shown in Table 6. As in each of these analyses, we show the baseline estimator (average of the training data) in column **A** for comparison. We also compare using ground truth building footprint area (column **B**) with CNN-detected building footprint area (column **C**) and CNN-detected building footprint area coupled with additional features extracted from the imagery data (column **D**). All of these approaches greatly outperform the baseline model (column **A**), while the magnitude of improvement in RMSE is generally greatest for using predictors composed of CNN-detected building footprint area coupled with additional features extracted from the imagery data (column **D**), which even outperforms the ground truth-based predictors in column **B** for aggregation levels of 200 × 200 m or larger. The additional features used in column **D** resulted in 14.9% RMSE improvements over the baseline for individual buildings compared to 7.3% for area alone. This is comparable to the 17.9% (column **B**) improvement that estimator based on the ground truth area boasts. Larger gains resulted from aggregating to 1000 × 1000 m regions which brought the improvement to 42.2% for the additional features of column **C**, up from 20.9% for using predicted area alone.

Lastly, we investigated estimating building energy consumption in San Diego. There are a total of 90 ZIP Codes for which we have data available in San Diego County. Its area is 11,722 square kilometers, therefore, these ZIP Codes are on average far larger than even the largest area of aggregation (1 km^2) that was used so far in these experiments. Additionally, since the area is larger, there is likely to be more heterogeneity in the data in terms of housing stock and energy consumption. For these experiments, we do not have data with both individual energy consumption and the specific location of the corresponding building, so we have to split up the building segmentation algorithm training and energy consumption estimation process across different datasets. For building segmentation training, we use a small dataset of 7,396 hand-labeled buildings (polygons of building footprints) in San Diego that was used to fine-tune the Gainesville building segmentation model trained on 798 buildings in Gainesville pretrained on the INRIA building segmentation dataset [87]. For energy consumption estimation we use the anonymized collection of 10,718 buildings in San Diego City with known energy consumption, building footprint area and ZIP Code location.

Having trained an individual building energy consumption regression model using the anonymized collection of buildings from 12 ZIP Codes, we test first on just San Diego City (18 ZIP Codes) with results shown in Table 7 column **A**, then both the city and all of the other San Diego County ZIP Codes (78 ZIP Codes) with results shown in column **B**. Compared to using the County average residential energy consumption for all buildings (baseline 1), our approach yielded a 34.1% improvement in RMSE at the ZIP Code level for San Diego City (column **A**) and are comparable against baseline 2 that uses a training set average for San Diego City – a 30.9% improvement (column **A**). For predicting the ZIP Code level energy consumption for San Diego County, we find a 6.4% improvement in RMSE across the 90 ZIP Codes from San Diego County (column **B**).

5.5. Results of the sensitivity analysis for quantity of training data needed for reliable output

Training data are hard to obtain for building energy consumption, so we explore the impact of the quantity of training data on the error in the energy consumption estimates. Over repeated experiments, we vary the size of the training dataset, holding the test set fixed, and evaluate performance using different samples of training data sampled with replacement. To ensure scales are matched on our axes across experiments we use R-squared as our metric for comparison: higher R-squared is more desirable. In each experiment, we are interested in where the plot appears to level off, saturating its performance. Alternatively, one could use it to approximate the number of examples needed to produce an estimator of a targeted performance.

For the Gainesville case, predicting energy consumption from CNN-derived area alone, we see in Fig. 3, unsurprisingly, the more data that are included, the better the performance. However, what is more interesting is that the performance increase from adding training begins producing marginal returns around 2,000 samples. Beyond that, very minor improvements are seen in performance. For larger areas of aggregation, even less data may be needed as the 1000 × 1000 m case appears to level off between 300 and 400 training samples. For individual buildings, 2,000 samples may be required, while for a square kilometer as few as 200 samples may be sufficient to approach near-peak performance. Additionally, over 95% of predictions made with at least 200 buildings in the sample outperformed the baseline average model in Gainesville.

For the San Diego sensitivity analysis results are shown in Fig. 4. In this case, individual building energy consumption performance over this larger region (an entire county) continues to increase with more data, which may be explained by the greater diversity in the larger region: the imagery datasets for Gainesville and San Diego County cover 315 km^2 and 9,131 km^2 respectively. Bear in mind that the training dataset comes from San Diego City which is unlikely to be representative of the entire County. Nevertheless, a dataset of as few as 300 buildings is enough to perform better than the baseline average estimator in San Diego City at the ZIP Code level. For San Diego County, new data steadily decreases the probability of the estimator underperforming the baseline. In fact,

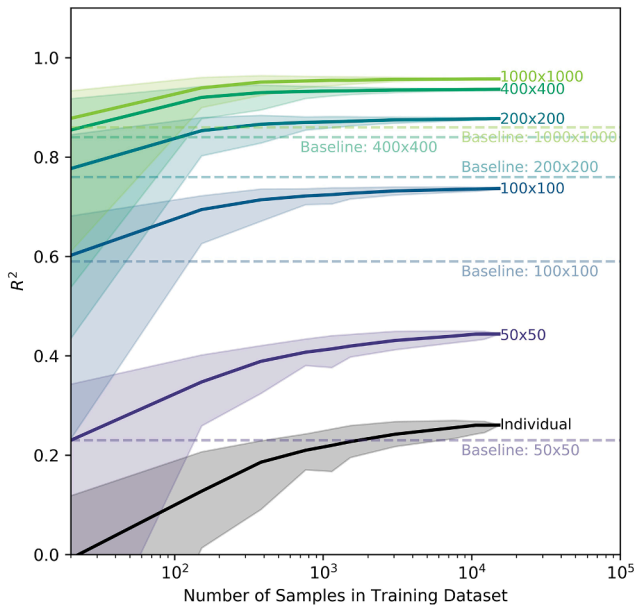

Fig. 3. Sensitivity analysis of the energy consumption estimation process varying the size of the training dataset for Gainesville, Florida over 100 repeated trials using a different subset of training data. Each line represents a median across trials for different levels of aggregation starting with the individual building level and increasing through 1 square km. Over the repeated runs of this model, the bands around each line show the range between percentiles 2.5% and 97.5%. The baseline for the case of individual building energy consumption prediction is an R^2 of 0.

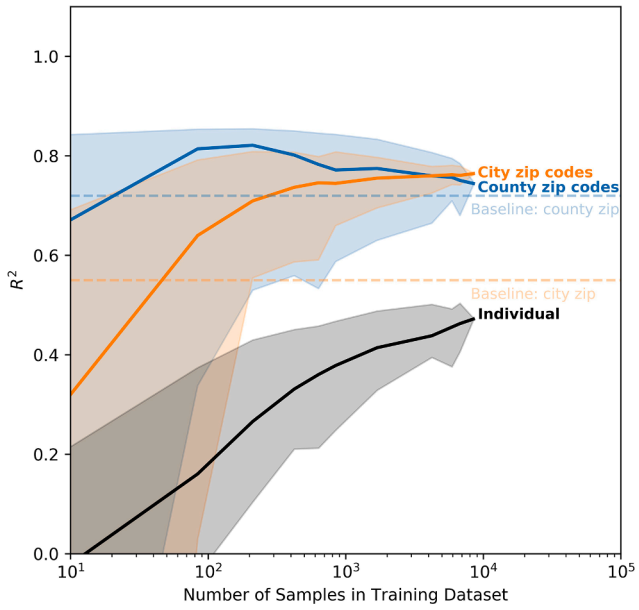

Fig. 4. Sensitivity analysis of the energy consumption estimation process varying the size of the training dataset for San Diego, California over 100 repeated trials where each trial uses a different subset of training data (sampled with replacement). The results show the median R^2 values for both individual building and ZIP Code level estimates. Baseline averages are shown for comparison. Over the repeated runs of this model, the bands around each line show the range between percentiles 2.5% and 97.5%.

even 100 buildings (0.01% of test set buildings in the County) is enough to produce a better-performing estimator in 75% of model runs, and for 4,000 buildings (0.56% of test set buildings in the County), this results in better-than-baseline performing estimates 90% of the time.

6. Conclusions

This work presents an end-to-end pipeline for residential building energy consumption based only on overhead imagery data. This method produces estimates that are generally lower-error compared to using the regional average building energy consumption. In Gainesville, our approach for detected buildings yielded individual buildings energy estimates with errors around 5,831 kWh per year (R^2 of 0.28), while the baseline (average) resulted in an error of 6,851 kWh per year. Aggregating the data over small "neighborhood" regions also resulted in additional improvements in building energy estimation accuracy. For a 1 square kilometer aggregation, the estimation error fell to between 3,784 and 4,754 kWh per year (R^2 between 0.81 and 0.97), while the baseline was 6,546 kWh per year. This meant our approach improved 7%-15% over baseline performance for estimating the energy consumption of individual residential buildings. By aggregating energy estimates up to 1 square kilometer neighborhoods, energy estimation improved to between 28% and 42% over the baseline. In San Diego County, while individual level predictions resulted in a R^2 of 0.38, aggregating up to the ZIP Code level demonstrated a 34% improvement in building energy estimation performance for San Diego City and a 6.4% improvement for San Diego County.

Features of the buildings included in the energy estimation model impacted performance, determining, in part, which end of the range of improvements over the baseline model estimates fell into. Extracting only the footprint area of each building was enough to produce the lower end of these performance improvement ranges. While area proved to be an essential feature for energy estimation prediction, additional features pushed the results towards the upper end of the performance range. Perimeter, neighborhood building density estimates, and other visual features extracted from imagery of each building and its surrounding using a deep learning model's encoder, pushed the performance to the upper end of the performance range. Another factor in performance was whether scoring metrics included occluded buildings (those not visible from above due to tree cover, etc.), with larger improvements resulting if visually occluded buildings were not factored into scoring. This demonstrates that these techniques perform better when the buildings are clearly visible. Quantitatively, the building detector correctly identifies 84% and 88% of buildings in Gainesville and San Diego, respectively, and the building type is classified with 99% accuracy for residential buildings and 74% accuracy for commercial buildings.

We also explored how the quantity of training data impacted the energy estimation performance achieved through a sensitivity analysis. For instance, in our Gainesville experiments, for individual building energy consumption estimates to achieve a 12% improvement over the baseline model around 2,000 samples of building data were needed, while for a 37% improvement in the 1 square kilometer aggregate estimates this requirement could be as low as 200–300 samples. These data requirements are not overly onerous and implies that this approach may be practical for applications in new and potentially larger geographies.

Overall, this approach demonstrates the feasibility of using overhead imagery to produce high-resolution estimates of residential building energy consumption that are generally more accurate than using what is typically available today: regional building energy consumption averages. This approach could be used to rapidly gain higher resolution insights on building energy consumption at a larger geographic scale than previously achievable without the need for large household surveys or proprietary data. This technique may enable researchers to better assess energy consumption trends around the world as residential housing development evolves and enhance evidence-based decision making for planning for energy system security and sustainability.

7. Limitations and future work

Among the limitations of this work, the availability of datasets with *both* known energy consumption and known location to provide adequate training data yielded a relatively small sample size for the number of buildings in our experiments. In Gainesville there was also substantial tree coverage in the imagery data for many buildings, limiting the efficacy of feature extraction and therefore performance for occluded structures. Another limitation was encountered in San Diego, where we had no buildings where both energy consumption and precise geolocation was known. This motivated our experimental choice of aggregating to the ZIP Code level. Lastly, our current feature extraction methods did not include techniques for directly estimating building height, which can enable improved estimates of overall building volume.

For future work, incorporating additional data could enhance this work in multiple ways. Gathering larger energy use datasets could enable measures of algorithm generalizability to be evaluated across more diverse regions. Other datasets may be added to the model pipeline such as data to provide information about building height such as Lidar data and street level imagery (similar to what is available from Google Street View). These data may have the potential to provide features on the physical characteristics of the dwelling (its height, building envelope material, window size etc.) and also potentially relevant socioeconomic data as well. Additionally, collecting more data on commercial buildings could allow similar work to be completed on estimating commercial building energy consumption. Lastly, if the number of buildings in a region is known, further work could be devoted to imputing the consumption of such buildings using both aggregate and individual estimates to improve regional energy consumption estimation.

CRediT authorship contribution statement

Artem Streltsov: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Jordan M. Malof:** Conceptualization, Methodology, Writing - review & editing. **Bohao Huang:** Methodology, Software. **Kyle Bradbury:** Conceptualization, Methodology, Validation, Resources, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge the work of the 2016 Data+ and 2016-2017 Bass Connections team at Duke University who laid the foundation of this work, identifying some of the datasets used in this study and putting together the first approach to the energy consumption estimation process. Those students are: Samit Sura, Hoël Wiesner, Min Chul (Mitchell) Kim, Jer Sheng (Sebastian) Lin, Jee Hye (Sophia) Park, Eric Peshkin, Nikhil Vanderklaauw, Yue (Joyce) Xi, Benjamin Brigman, and Sunith Suresh. We would also like to thank Leslie Collins, Timothy Johnson, and Richard Newell for ideas and suggestions around this project as well as Will Niver for his assistance with the literature review and the introduction for this work and Wei (Wayne) Hu for assistance with finalizing some of the figures in this work. This work was supported in part by National Science Foundation Grant no. OIA-1937137.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apenergy.2020.116018.

References

[1] International Energy Agency, "Transition to sustainable buildings: strategies and opportunities to 2050," International Energy Agency, 2013. [Online]. Available: http://www.iea.org/Textbase/npsum/building2013SUM.pdf.
[2] Cao X, Dai X, Liu J. Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. Energy Build Sep. 2016;128:198–213.
[3] Li W, et al. Modeling urban building energy use: A review of modeling approaches and procedures. Energy Dec. 2017;141:2445–57. https://doi.org/10.1016/j.energy.2017.11.071.
[4] Reinhart CF, Cerezo Davila C. Urban building energy modeling – A review of a nascent field. Build Environ 2016;97:196–202. https://doi.org/10.1016/j.buildenv.2015.12.001.
[5] Hirsh RF, Koomey JG. Electricity consumption and economic growth: a new relationship with significant consequences? Electr J Nov. 2015;28(9):72–84. https://doi.org/10.1016/j.tej.2015.10.002.
[6] Morna Isaac, van Vuuren Detlef. Modeling global residential sector energy demand for heating and air conditioning in the context of climate change. Energy Policy Feb. 2009;37(2):507–21.
[7] "Decision adopting rules to provide access to energy usage and usage-related data while protecting privacy of personal data." May 05, 2014, Accessed: May 20, 2020. [Online]. Available: https://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M090/K845/90845985.PDF.
[8] "Residential Energy Consumption Survey (RECS) - Data - U.S. Energy Information Administration (EIA)." https://www.eia.gov/consumption/residential/data/2015/ (accessed Jun. 04, 2020).
[9] "Dataport," Pecan Street Inc. https://www.pecanstreet.org/dataport/ (accessed Jun. 04, 2020).
[10] "Gainesville-Green.com." http://gainesville-green.com/.
[11] "REDD," Reference Energy Disaggregation Data Set. http://redd.csail.mit.edu/ (accessed Jun. 04, 2020).
[12] "NILM@CMU," BLUED: Building-Level fUlly labeled Electricity Disaggregation dataset. http://portoalegre.andrew.cmu.edu:88/BLUED/ (accessed Jun. 04, 2020).
[13] "PLAID," A Public Dataset of High Resolution for Load Identification Research. http://www.plaidplug.com/ (accessed Jun. 04, 2020).
[14] "GREEND," An energy dataset of households in Austria and Italy. https://sourceforge.net/projects/greend/ (accessed Jun. 04, 2020).
[15] Makonin S. AMPds: Almanac of minutely power dataset (R2013). Harvard Dataverse 2015. https://doi.org/10.7910/DVN/MXB7VO.
[16] Reinhardt A. areinhardt/tracebase; 2020.
[17] "UK Dale 2017," UK Domestic Appliance Level Electricity (UK-DALE-2017)-Disaggregated appliance/whole house power dataset. https://data.ukedc.rl.ac.uk/browse/edc/efficiency/residential/EnergyConsumption/Domestic/UK-DALE-2017/UK-DALE-FULL-disaggregated (accessed Jun. 04, 2020).
[18] Foucquier A, Robert S, Suard F, Stéphan L, Jay A. State of the art in building modelling and energy performances prediction: A review. Renew Sustain Energy Rev Jul. 2013;23:272–88. https://doi.org/10.1016/j.rser.2013.03.004.
[19] Clarke J. Energy simulation in building design. 2nd ed. 2007.
[20] Underwood C, Yik F. Modelling methods for energy in buildings. John Wiley & Sons, Ltd; 2008.
[21] Crawley DB, et al. EnergyPlus: creating a new-generation building energy simulation program. Energy Build 2001;33(4):319–31. https://doi.org/10.1016/S0378-7788(00)00114-6.
[22] Qin R, Yan D, Zhou X, Jiang Y. Research on a dynamic simulation method of atrium thermal environment based on neural network. Build Environ 2012;50:214–20. https://doi.org/10.1016/j.buildenv.2011.11.001.
[23] Zhai Z, Chen Q, Haves P, Klems JH. On approaches to couple energy simulation and computational fluid dynamics programs. Build Environ 2002;37(8):857–64. https://doi.org/10.1016/S0360-1323(02)00054-9.
[24] Barbason M, Reiter S. Coupling building energy simulation and computational fluid dynamics: Application to a two-storey house in a temperate climate. Build Environ 2014;75:30–9. https://doi.org/10.1016/j.buildenv.2014.01.012.
[25] Musy M, Winkelmann F, Wurtz E, Sergent A. Automatically generated zonal models for building air flow simulation: principles and applications. Build Environ 2002;37(8):873–81. https://doi.org/10.1016/S0360-1323(02)00050-1.
[26] Tittelein P, Wurtz E, Achard G. Simspark platform evolution for lowenergy building simulation. Int Sci J Altern Energy Ecol 2008;62.
[27] Boyer H, Chabriat JP, Grondin-Perez B, Tourrand C, Brau J. Thermal building simulation and computer generation of nodal models. Build Environ 1996;31(3):207–14. https://doi.org/10.1016/0360-1323(96)00001-7.
[28] (John) Zhai Z, Johnson M-H, Krarti M. Assessment of natural and hybrid ventilation models in whole-building energy simulations. Energy Build., vol. 43, no. 9, pp. 2251–2261, Sep. 2011, doi: 10.1016/j.enbuild.2011.06.026.
[29] Olsen EL, (Yan) Chen Q. Energy consumption and comfort analysis for different low-energy cooling systems in a mild climate. Energy Build 2003;35(6):560–71. https://doi.org/10.1016/S0378-7788(02)00164-0.
[30] Phan L, Lin C-X. A multi-zone building energy simulation of a data center model with hot and cold aisles. Energy Build 2014;77:364–76. https://doi.org/10.1016/j.enbuild.2014.03.060.
[31] Yang X, Zhao L, Bruse M, Meng Q. An integrated simulation method for building energy performance assessment in urban environments. Energy Build 2012;54:243–51. https://doi.org/10.1016/j.enbuild.2012.07.042.

[32] Dols WS, Emmerich SJ, Polidoro BJ. Coupling the multizone airflow and contaminant transport software CONTAM with EnergyPlus using co-simulation. Build Simul 2016;9(4):469–79. https://doi.org/10.1007/s12273-016-0279-2.

[33] Swan LG, Ugursal VI. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renew Sustain Energy Rev 2009;13(8): 1819–35. https://doi.org/10.1016/j.rser.2008.09.033.

[34] Hong S, Paterson GA, Burman E, Steadman P, Mumovic D. A comparative study of benchmarking approaches for non-domestic buildings : Part 1 – Top-down approach. Int J Sustain Built Environ 2013;2(2):119–30.

[35] Li C, Ding Z, Zhao D, Yi J, Zhang G. Building energy consumption prediction: an extreme deep learning approach. Energies 10(10), Art. no. 10, Oct. 2017, doi: 10.3390/en10101525.

[36] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy Build May 2005;37(5):545–53. https://doi.org/10.1016/j.enbuild.2004.09.009.

[37] Huebner GM, Hamilton I, Chalabi Z, Shipworth D, Oreszczyn T. Explaining domestic energy consumption – The comparative contribution of building factors, socio-demographics, behaviours and attitudes. Appl Energy 2015;159:589–600. https://doi.org/10.1016/j.apenergy.2015.09.028.

[38] Ma J, Cheng JCP. Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology. Appl Energy 2016;183:182–92. https://doi.org/10.1016/j.apenergy.2016.08.079.

[39] Min J, Hausfather Z, Lin QF. A high-resolution statistical model of residential energy end use characteristics for the United States. J Ind Ecol, 14(5), pp. 791–807, doi: 10.1111/j.1530-9290.2010.00279.x.

[40] Chen J, Wang X, Steemers K. A statistical analysis of a residential energy consumption survey study in Hangzhou, China. Energy Build 2013;66:193–202. https://doi.org/10.1016/j.enbuild.2013.07.045.

[41] Mastrucci A, Baume O, Stazi F, Leopold U. Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam. Energy Build 2014;75:358–67. https://doi.org/10.1016/j.enbuild.2014.02.032.

[42] Kelly Scott. Do homes that are more energy efficient consume less energy?: A structural equation model of the English residential sector. Energy 2011;36(9): 5610–20.

[43] Fikru MG, Gautier L. The impact of weather variation on energy consumption in residential houses. Appl Energy 2015;144:19–30. https://doi.org/10.1016/j.apenergy.2015.01.040.

[44] Sanquist TF, Orr H, Shui B, Bittner AC. Lifestyle factors in U.S. residential electricity consumption. Energy Policy 2012;42:354–64. https://doi.org/10.1016/j.enpol.2011.11.092.

[45] Howard B, Parshall L, Thompson J, Hammer S, Dickinson J, Modi V. Spatial distribution of urban building energy consumption by end use. Energy Build 2012; 45:141–51. https://doi.org/10.1016/j.enbuild.2011.10.061.

[46] Kavousian A, Rajagopal R, Fischer M. Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. Energy 2013;55:184–94. https://doi.org/10.1016/j.energy.2013.03.086.

[47] Kadir Amasyali NE-G. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev, 81:2018;1192–205, doi: 10.1016/j.rser.2017.04.095.

[48] Ahmad T, Chen H, Guo Y, Wang J. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. Energy Build 2018;165:301–20. https://doi.org/10.1016/j.enbuild.2018.01.017.

[49] Streltsov A, Bradbury K, Malof J. Automated building energy consumption estimation from aerial imagery. In: IGARSS 2018–2018 IEEE international geoscience and remote sensing symposium; 2018. p. 1676–9.

[50] Min B, Gaba KM, Sarr OF, Agalassou A. Detection of rural electrification in Africa using DMSP-OLS night lights imagery. Int J Remote Sens 2013;34(22):8118–41. https://doi.org/10.1080/01431161.2013.833358.

[51] Dugoua E, Kennedy R, Urpelainen J. Satellite data for the social sciences: measuring rural electrification with night-time lights. Int J Remote Sens 2018;39 (9):2690–701. https://doi.org/10.1080/01431161.2017.1420936.

[52] Letu H, et al. Estimating energy consumption from night-time DMPS/OLS imagery after correcting for saturation effects. Int J Remote Sens 2010;31(16):4443–58. https://doi.org/10.1080/01431160903277464.

[53] Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S. Combining satellite imagery and machine learning to predict poverty. Science 2016;353(6301):790–4. https://doi.org/10.1126/science.aaf7894.

[54] Malof JM, Collins LM, Bradbury K, Newell RG. A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in aerial imagery. In: 2016 IEEE international conference on renewable energy research and applications (ICRERA), Nov. 2016, pp. 650–654, doi: 10.1109/ICRERA.2016.7884415.

[55] Yu J, Wang Z, Majumdar A, Rajagopal R. DeepSolar: A machine learning framework to efficiently construct a solar deployment database in the United States. Joule 2018;2(12):2605–17. https://doi.org/10.1016/j.joule.2018.11.021.

[56] Huang B, et al. Large-scale semantic classification: outcome of the first year of inria aerial image labeling benchmark. In: IGARSS 2018 - 2018 IEEE international geoscience and remote sensing symposium, Valencia, Jul. 2018, pp. 6947–6950, doi: 10.1109/IGARSS.2018.8518525.

[57] Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H, 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. ArXiv180202611 Cs, Aug. 2018, Accessed: Jun. 04, 2020. [Online]. Available: http://arxiv.org/abs/1802.02611.

[58] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Presented at the international conference on medical image computing and computer-assisted intervention, May 2015, Accessed: Jan. 05, 2018. [Online]. Available: http://arxiv.org/abs/1505.04597.

[59] Zhou L, Zhang C, Wu M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In: 2018 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), Salt Lake City, UT, USA, Jun. 2018. p. 192–1924, doi: 10.1109/CVPRW.2018.00034.

[60] Demir I, et al. DeepGlobe 2018: A challenge to parse the earth through satellite images. 2018 IEEECVF Conf. Comput. Vis. Pattern Recognit. Workshop CVPRW, pp. 172–17209, Jun. 2018, doi: 10.1109/CVPRW.2018.00031.

[61] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Presented at the 2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, USA, Dec. 2015, doi: 10.1109/CVPR.2016.90.

[62] Maggiori E, Tarabalka Y, Charpiat G, Alliez P. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In: 2017 IEEE international geoscience and remote sensing symposium (IGARSS), Fort Worth, TX, Jul. 2017, pp. 3226–3229, doi: 10.1109/IGARSS.2017.8127684.

[63] Haas R, Biermayr P, Zoechling J, Auer H. Impacts on electricity consumption of household appliances in Austria: A comparison of time series and cross-section analyses. Energy Policy 1998;26(13):1031–40. https://doi.org/10.1016/S0301-4215(98)00057-3.

[64] Ratio of household electricity use to national electricity consumption by country. Center For Global Development. https://www.cgdev.org/media/ratio-household-electricity-use-national-electricity-consumption-country (accessed Dec. 18, 2019).

[65] Robinson C, et al. Machine learning approaches for estimating commercial building energy consumption. Appl Energy 2017;208:889–904. https://doi.org/10.1016/j.apenergy.2017.09.060.

[66] Amiri SS, Mottahedi M, Asadi S. Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the U.S. Energy Build 2015;109:209–16. https://doi.org/10.1016/j.enbuild.2015.09.073.

[67] Rahman A, Srikumar V, Smith AD. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl Energy 2018;212:372–85. https://doi.org/10.1016/j.apenergy.2017.12.051.

[68] astr93, astr93/pyimannotate. 2020.

[69] "U.S. Geological Survey, 2016, USGS earth resources observations and science center (EROS)," https://earthexplorer.usgs.gov/. .

[70] Alachua county property appraiser, "BuildingFootprints." http://data-acpa.opendata.arcgis.com/datasets/80deed550f0a4aa985a475deb8d9b4f1_0.

[71] "Energy Data Access." https://energydata.sdge.com/ (accessed Dec. 17, 2018).

[72] "Building Performance Database | Department of Energy." https://www.energy.gov/eere/buildings/building-performance-database (accessed Dec. 17, 2018).

[73] "City of San Diego Open Data Portal," City of San Diego Open Data Portal. /datasets/zoning/ (accessed Dec. 18, 2018).

[74] Ohleyer S. Building segmentation on satellite images. ENS Paris-Saclay, p. 5.

[75] Haralick RM, Shapiro LG. Image segmentation techniques. Comput Vis Graph Image Process Jan. 1985;29(1):100–32. https://doi.org/10.1016/S0734-189X(85)90153-7.

[76] Cordts M, et al. The cityscapes dataset for semantic urban scene understanding. ArXiv160401685 Cs, Apr. 2016, Accessed: Jun. 05, 2020. [Online]. Available: http://arxiv.org/abs/1604.01685.

[77] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks?. ArXiv14111792 Cs, Nov. 2014, Accessed: Jun. 05, 2020. [Online]. Available: http://arxiv.org/abs/1411.1792.

[78] Camilo J, Collins L, Bradbury K, Malof J. Application of a semantic segmentation convolutional neural network for accurate automatic detection and mapping of solar photovoltaic arrays in aerial imagery. DC: Present. IEEE Appl. Imag. Pattern Recognit. Workshop Wash; 2017.

[79] Douglas DH, Peucker TK. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. In: Classics in cartography, John Wiley & Sons, Ltd; 2011. p. 15–28.

[80] Ürge-Vorsatz D, Cabeza LF, Serrano S, Barreneche C, Petrichenko K. Heating and cooling energy trends and drivers in buildings. Renew Sustain Energy Rev 2015;41: 85–98. https://doi.org/10.1016/j.rser.2014.08.039.

[81] Deng J, Dong W, Socher R, Li L, Li Kai, Fei-Fei Li. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition 2009:248–55. https://doi.org/10.1109/CVPR.2009.5206848.

[82] Wold S, Esbensen K, Geladi P. Principal component analysis, p. 16.

[83] Chelikani S, Collins L, Bradbury K, Malof J. Trading spatial resolution for improved detection accuracy in remote sensing imagery: an empirical study using synthetic data. DC: Present. IEEE Appl. Imag. Pattern Recognit. Workshop Wash; 2017.

[84] "Home Page-California Energy Commission." https://www.energy.ca.gov/ (accessed Sep. 13, 2019).

[85] "Asheville, North Carolina: Reducing Electricity Demand through Building Programs & Policies," U.S. Department of Energy. [Online]. Available: https://www.energy.gov/sites/prod/files/2017/10/f37/Cities-LEAP_Data%20to%20Decisions_Asheville%2C%20North%20Carolina.pdf.

[86] Residential electricity consumption. CSE, Jul. 02, 2018. https://sites.energycenter.org/equinox/dashboard/residential-electricity-consumption (accessed Jul. 24, 2019).

[87] Inria Aerial Image Labeling Dataset. https://project.inria.fr/aerialimagelabeling/ (accessed Oct. 03, 2019).