

# TODOS: Thermal sensOr Data-driven Occupancy Estimation System for Smart Buildings

Hamid Rajabi, Xianzhong Ding, Wan Du, and Alberto Cerpa  
{hrajabi2,xding5,wdu3,acerpa}@ucmerced.edu  
University of California, Merced

## ABSTRACT

Occupancy sensing and estimation in large commercial buildings has become a significant problem to be solved, with applications ranging from occupancy-based HVAC control to space planning, and security, etc. Thermal sensing is a promising technology to solve this problem, being easy to deploy in practice and allowing an actual occupancy count in a particular room without violating the data and privacy concerns. While initial strides have been made to solve this problem with thermal arrays, there are many problems that remain unsolved, including accuracy performance, overlapping of sensing areas that lead to under/over-counting, and data training requirements for different zones.

In this paper, we introduce TODOS<sup>1</sup>, a novel system for estimating occupancy in intelligent buildings. TODOS uses a low-cost, low-power thermal sensor array along with a passive infrared sensor. We introduce a novel data processing pipeline that allows us to automatically extract features from the thermal images using an artificial neural network. Through an extensive experimental evaluation<sup>2</sup>, we show that TODOS provides occupancy detection accuracy of 98% to 100% under different scenarios. In addition, it solves the issue of occupancy over/under-counting by overlapping sensing areas when using multiple thermal sensors in large rooms. This is done by treating the entire area as a single input thermal image instead of partitioning the area into multiple thermal images individually processed. Furthermore, TODOS introduces a data augmentation technique that allows the generation of training data for rooms of different sizes and shapes, without requiring specific training data from each room. Using these data, TODOS can train specifically designed neural networks optimized for any room size and shape, and achieve almost the same level of occupancy detection accuracy in rooms where experimental labeled training data is available, making it a viable solution that generalizes to the different rooms in large buildings.

## CCS CONCEPTS

• **Computer systems organization** → **Sensor networks**; • **Computing methodologies** → **Neural networks**; *Image and video acquisition*; **Supervised learning by classification**.

<sup>1</sup>TODOS, from Spanish meaning *everybody*

<sup>2</sup>Full data experimental traces can be found online: <https://doi.org/10.21227/y3dv-8b19>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*BuildSys '23, November 15–16, 2023, Istanbul, Turkey*

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0230-3/23/11.

<https://doi.org/10.1145/3600100.3623753>

## KEYWORDS

smart buildings, occupancy estimation, HVAC systems, thermal occupancy sensor, neural networks, data augmentation

## ACM Reference Format:

Hamid Rajabi, Xianzhong Ding, Wan Du, and Alberto Cerpa. 2023. TODOS: Thermal sensOr Data-driven Occupancy Estimation System for Smart Buildings. In *The 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '23)*, November 15–16, 2023, Istanbul, Turkey. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3600100.3623753>

## 1 INTRODUCTION

We spend 87% of our time inside buildings [13, 14, 22], making buildings an important part of our lives. In particular, in large commercial buildings, it is important to know at any time what is the current and predictive occupancy matrix of all the zones/rooms in a building. Having detailed occupancy information about the number of people in real-time and in the near future for each zone allows for control of the HVAC systems in a much more efficient manner [3, 5, 15], conditioning zones for temperature only when occupied, and adjusting ventilation rates based on the actual number of occupants. Occupancy data can also aid with many building controls and management decisions. The work in [1] identified some of the important factors leading to poor maintenance strategies in buildings such as a lack of understanding of occupant satisfaction, inadequate staffing levels, and inefficient maintenance delivery. To enhance maintenance management operations, Moretti et al. [26] used occupancy data to alert the maintenance staff for cleaning operations when a threshold (defined based on the number of occupants using a facility) was reached. Facility managers can leverage occupancy sensing technologies to determine space usage and occupancy mobility patterns for security in buildings [2, 9], improve health and safety by adjusting ventilation rate for disease transmission according to the number of occupants [36] and locate individuals in case of an emergency evacuation by first responders [37].

Among the many sensing modalities available, thermal occupancy techniques [6, 8, 35, 38] seem the most promising for the task at hand. This is because they are non-intrusive, low-cost, low-power, easy to deploy, and perhaps more importantly, they can provide accurate occupancy counts without violating data and privacy concerns to the building's occupants [16]. While thermal occupancy technology seems very promising, there are still several issues that prevent full adoption. First, the features extracted from thermal images have been manually derived, but they may still not be powerful enough to provide a higher level of accuracy required for many applications [13]. Second, most of the studies carried out have concentrated on processing the input of a single sensor. When deploying multiple thermal sensors in a single zone/room, a

small amount of sensing overlapping is unavoidable to provide full coverage. This leads to issues of occupancy over/under-counting, when multiple sensors deployed in the same zone count the occupant in an overlapping area twice (over-counting) or when the portion of the occupant thermal signature is not sufficient to lead to a count in any of the sensors (under-count). Finally, data-driven processing pipelines require a significant amount of labeled training data to build the models. The input of these models for any specific zone/room will depend on the distribution and location of the sensors being deployed, which roughly coincide with the size and shape of the room. Consequently, since the size and shape of the rooms in a building are different for many of them, this means that a significant effort must be carried out to obtain adequate training data for each type of room (size and shape). This last point is a significant roadblock to the wide adoption of this promising technology in practice.

In this paper, we introduce TODOS, a novel system for estimating occupancy in intelligent buildings. TODOS uses a low-cost, low-power thermal sensor array along with a passive infrared sensor. We introduce a novel data processing pipeline that allows us to automatically extract features from the thermal images using a convolutional neural network. Through an extensive experimental evaluation done in multiple rooms at multiple university buildings, we show that TODOS provides occupancy detection accuracy of 98% to 100% under different scenarios. In addition, it solves the issue of occupancy over/under-counting by overlapping sensing areas when using multiple thermal sensors in large rooms. This is done by treating the entire area as a single input image instead of partitioning the area into multiple thermal images individually processed. Moreover, TODOS introduces a data augmentation technique that allows the generation of training data for rooms of different sizes and shapes, without requiring specific training data from each room. Using these data, TODOS can train specifically designed neural networks optimized for any room size and shape, and achieve almost the same level of occupancy detection accuracy in rooms where experimental labeled training data is available, making it a viable solution that generalizes to the different rooms in large buildings. The main contributions of this work are as follows:

- We developed TODOS, a thermal-based occupancy sensing system that uses low-power, low-cost, easy-to-deploy sensors and a novel processing pipeline that automatically extracts thermal image features, treats all the sensor data as a global thermal image per zone/room and achieves excellent accuracy in occupancy detection.
- We provided more robust training data sets using data augmentation techniques for different combinations of sensors. This allows TODOS to have better training, even for cases that were unseen in the original experimental data gathering. We then replicated this augmented data set and concatenated them according to the room size and its geometrical shape, so we can train our model for rooms of any shape and size.
- We run an extensive experimental campaign, to see how TODOS generalizes to different buildings, air diffusers, and rooms with different sizes and shapes, and compared TODOS's performance using both local training labeled data as well as augmented data from a different room, showing that TODOS can achieve excellent performance without the need to collect local data.

- We performed Energy+ simulations to show the energy and quality of comfort impact that a more accurate occupancy estimation may have when using occupancy-based HVAC control.

## 2 RELATED WORK

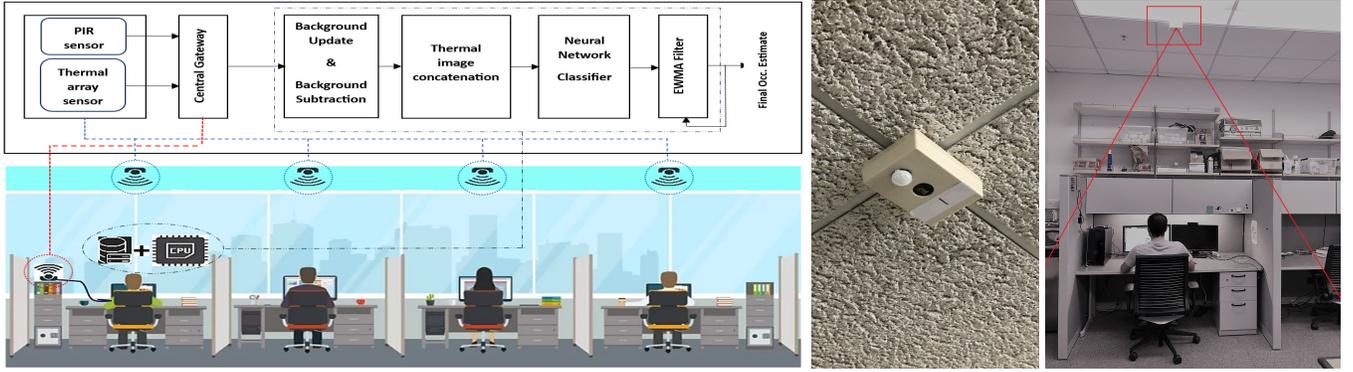
Occupancy estimation schemes can be broadly categorized into user-based and user-free schemes. In user-based occupancy sensing systems such as [10, 11, 18], the building users carry a device or tag and are capable of delivering acceptable performance ranging from 83% to 94% accuracy. However, such systems suffer from degraded performance issues when the users fail to carry the device. In addition, many of them suffer from regulatory barriers to entry due to data and privacy requirement regulations [16]. User-free schemes [4, 6, 27, 31, 40] do not suffer from the problems mentioned above. There are many different sensing technologies, including video-based sensing systems [7, 21, 34],  $CO_2$  sensors [20, 28, 40], vibration systems [25, 29, 30], WiFi-based systems [24, 39], and thermal-based occupancy sensing modalities [6, 27, 38].

The critical feature of why thermal-based sensing works is the temperature differential between occupants and their indoor environment, which is usually much lower when it is thermally conditioned by an HVAC system. These systems [6, 8, 35, 38] are capable of detecting occupants in large groups both in static (e.g. sitting) and dynamic (e.g. moving) positions. They do require LOS for detection, but differently from camera-based systems, they are significantly less intrusive, since they just measure temperature instead of recording a full video [16]. This means that the technology is very suitable for deployments in indoor environments while complying with data and privacy requirements.

The authors in [8] proposed an accurate privacy-preserved thermal occupancy sensing system. However, this system design lacks an energy-saving mode in the sensor hardware when it is idle. In addition, they use very similar processing techniques to ThermoSense [6], including very similar features that are fed to different types of classifiers to do occupancy estimation. In TODOS, and similarly to ThermoSense [6], we do include an additional PIR sensor that allows us to duty-cycle the thermal array when the sensing area is empty. Moreover, we propose to use a deeper neural network technique optimized for occupancy sensing and a technique to generalize to rooms/zones of different sizes and shapes that significantly reduces the requirements for training and testing data in different buildings.

The closest related work to TODOS is ThermoSense [6]. The node hardware design is very similar. ThermoSense processing pipeline includes active pixel detection, background subtraction, and connected components as a preprocessing stage. It then extracts the total number of active pixels passing a threshold, the number of the blobs (from connected components), and the size of the largest blob. These features are then used as inputs for different types of classifiers (including linear regression, KNN, and ANN), to classify the number of occupants in the sensing area covered by the sensor. In their evaluation, they use the linear classifier since the performance differences among the different classifiers evaluated were not significant and it was easier to implement the linear classifier in computationally limited TelosB motes hardware.

Tyndall et al. [35] presented a thermal-based system that is built upon the work in ThermoSense (i.e. combination of PIR and thermal



(a) TODOS processing pipeline. The data used for model training depends on the shape and size of the room the sensor(s) are deployed. (b) Thermal sensor node (c) The area covered by a sensor mote with an occupant.

Figure 1: TODOS Deployment and System Overview

array sensor), but with a different heat map array size and different classifiers such as MLP, SVM, Naive-Bayes, KNN, and linear regression processing pipeline. The reported performance is similar to the one reported by ThermoSense.

In our work with TODOS, we do a similar preprocessing stage to ThermoSense, including active pixel detection and background subtraction. However, this information is fed into a deep convolutional neural network (CNN) which is significantly deeper than the ANN used in ThermoSense, and we let the CNN learn and extract the features of the thermal images instead of manually setting the input features. Moreover, we solved the problem of over/under-counting occupants when different sensors detect an occupant at the edge of their sensing range, by aggregating all the sensor data in one big zone image. Both ThermoSense and TODOS use at the end of their respective processing pipelines an EWMA low-pass filter to get rid of spurious occupancy readings over time. Both [6, 35] are used in our performance evaluation for comparison with the state-of-the-art schemes available in the literature. As a thorough comparison, we also tested and analyzed TODOS's performance under two different standard CNN architectures.

### 3 TODOS OVERVIEW

Fig. 1a shows TODOS's processing pipeline. The sensor nodes are deployed in the ceiling of a particular room, with a particular deployment of sensors depending on the size and shape of the room. All the nodes sense a specific part of the room, sample every 20 seconds both the PIR and thermal array sensors, and transmit the thermal image using an 802.15.4 Zigbee radio to a central gateway (a TelosB mote attached to the USB port of a CPU). The gateway, then, performs a radio-to-serial conversion and transfers thermal images to the processing server. Within the processing server, the images are concatenated based on a preconfigured order of images (that depends on the node id). This new larger thermal image is fed to a convolutional neural network doing classification on the number of occupants. Finally, this output is passed through a low-pass EWMA filter to remove high-frequency noise that may appear from time to time and improve the overall estimate.

#### 3.1 Hardware Sensor Components

TODOS sensing hardware consists of a PIR sensor and a thermal sensor array connected together to a TelosB sensor mote (see Fig. 1b).

The PIR delivers a binary indication of occupancy, and it allows for saving energy when no activity is detected. The PIR has a detection viewing angle of  $102^\circ \times 92^\circ$ , horizontally by vertically. It is able to detect motion up to 12m. The other sensor is a thermal array sensor to determine the number of occupants underneath (see Fig. 1c). The sensor is a Panasonic Grid-Eye thermal array, which consists of an  $8 \times 8$  thermal sensor that can cover an area of approximately  $2.5m \times 2.5m$  with a ceiling height of 3m. It is capable of measuring 64 temperature values, with temperatures ranging from  $-20^\circ C$  to  $80^\circ C$  with an accuracy of  $\pm 2.5^\circ C$ .

#### 3.2 Thermal Background

A background subtraction between a new thermal image (with one or more occupants) and the thermal background when there were no occupants is done to infer occupancy. This operation works because the temperature of the background objects (e.g. chairs, desks, floor, etc.) tends to be around the conditioning temperature of the room, which usually oscillates between  $22^\circ C$  and  $26^\circ C$  for HVAC-conditioned spaces in commercial buildings. Since the temperature of a human head and limbs is around  $37^\circ C$ , the difference is significant enough to work for the detection of occupants.

A thermal background map is maintained to distinguish between warm objects like computers or refrigerators and occupants. If the PIR has detected no movement for a certain period of time (e.g. 10-15 minutes), the background gets updated and the standard deviation is also updated for each grid component. Note that if an occupant remains in a space for a quite long period of time, it is possible that the background changes during this period. To adjust the background in this case, a few grid points are chosen with the lowest temperatures as the scaling factor. The points with the lowest temperatures are most likely unoccupied and can be used to update the old background. Those scale points are then divided by the old background and averaged to find a multiplier to update the previous background.

We perform background subtraction by using the difference between the background and the current thermal map and applying a standard deviation-based threshold to this difference, we can create an  $8 \times 8$  binary matrix representing the significantly warm points from the thermal map.

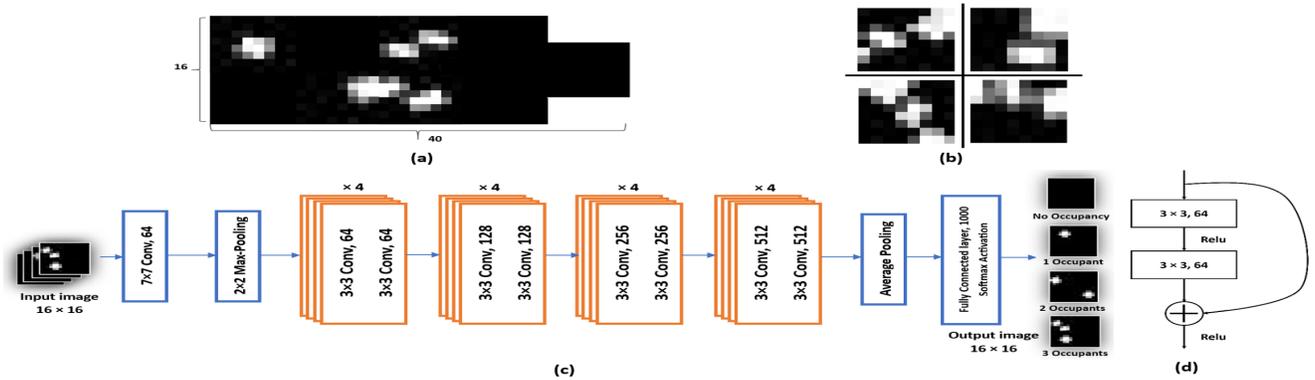


Figure 2: (a) An example of a concatenated thermal image of an entire room ( $40 \times 16$  pixels), (b) some transformation of the  $8 \times 8$  array heat map on the occupancy level 3 due to data augmentation, (c) ResNet-18 architecture used in TODOS for a room with four  $8 \times 8$  sensors forming input and output images of  $16 \times 16$ , and (d) a typical residual function used in ResNet-18.

### 3.3 Pre-processing Data

After reception of all the  $8 \times 8$  thermal images, they are concatenated using a preconfigured order based on the placement of the sensors in the room. For example, for a square room with 4 sensors, the concatenated input image will be  $16 \times 16$  pixels (using node ID to find the correct placement). In other words, the final concatenated thermal image is a thermal representation of the entire room/zone. We considered two different sets of images to input to the classifier. Each pixel image could be either in (1) a binary format, which is the result of the background subtraction and thresholding process explained in § 3.2, or (2) a grayscale of 8 bits (i.e. 256 values), that is obtained through the background subtraction but not thresholded. Fig. 2a shows an example of a concatenated grayscale thermal image ( $40 \times 16$  pixels) representing the entire room, assembled from nine (9) individual thermal sensor images ( $8 \times 8$  pixels). Both types of input data (i.e. binary and grayscale images) get processed through the deep learning module to evaluate the performance of each.

### 3.4 Neural Network Classifier

Fig. 2c shows the convolutional neural network (CNN) architecture used in our paper, which is Residual Network-18 (ResNet-18) [19] capable of improving the efficiency of CNN while minimizing the errors. In general, ResNet family architectures start with a convolutional layer of 64 kernels with size  $7 \times 7$  i.e. ( $7 \times 7, 64$ ), followed by a  $3 \times 3$  max-pooling layer. In ResNet-18, it is then followed by 16 convolutional layers of: a ( $3 \times 3, 64$ ) layer, another ( $3 \times 3, 128$ ) layer, then a ( $3 \times 3, 256$ ) layer, and a fourth layer of ( $3 \times 3, 512$ ), and each layer is repeated 4 times. At the end of the ResNet architecture, there is an average pooling layer followed by a fully connected layer with 1000 nodes using the Softmax activation function. ResNet is made up of a series of residual blocks with skip connections to address the vanishing gradient problem. Fig. 2d shows a typical residual function in ResNet-18.

### 3.5 EWMA Filter

While the classifier provides a good estimate of the occupancy count on a single thermal image, it is a memoryless system, in the

Table 1: Cross-validation parameters used in TODOS.

Parameter	Values Tested	Best Values
Epoch	35, 30, 15, 10, 5	35
Batch Size	5, 10, 20, 64	10
Learning Rate	0.01, 0.001, 0.0001	0.001

sense that its output is not affected by previous inputs. In our application, we sample occupancy every 20 seconds. However, human occupancy over time has specific time patterns that can be leveraged in order to improve accuracy. For example, it is rare for the occupancy of a room to be 8 people, and then be 1, and then 8 again, and so forth. So spurious values of a classifier could be improved by applying a filter. To get a smooth final occupancy estimate and remove any spurious values over time, we apply an Exponentially Weighted Moving Average (EWMA) low-pass filter. The filter is defined as:

$$y(t) = \alpha \hat{y}(t) + (1 - \alpha)y(t - 1) \quad (1)$$

where  $y(t)$  is the current occupancy estimate,  $\hat{y}(t)$  is the current occupancy sample from our classifier at time  $t$  and  $y(t - 1)$  is the previous occupancy at  $t - 1$ .  $\alpha$  is a real value between 0 and 1.

### 3.6 Data Augmentation and Model Training

The dataset in a neural network model needs to be rich and sufficient, so the model performs better and more accurately. However, collecting and labeling that rich data can be a costly process. Transformations in datasets via Data Augmentation techniques allow us to reduce those operational costs [23]. In this work, we have used width shift, height shift, shearing, zooming, rotation, horizontal flip, and vertical flip geometric augmentations. These types of geometric transformations are safe in our case since they represent occupants in different positions with respect to the sensor position, and for different ceiling heights that can be found in the different rooms in a building. Fig. 2b shows some examples of changes we applied to our existing dataset. In our work, we used an augmented dataset containing 7000 to 10000 thermal images for *each* occupancy case between 0 and 11.

We partition our data into training (80%) and testing (20%) sets. The data used for training is of the form  $\{\vec{x}_N, y_n\}_{n=1}^N$  with  $\vec{x}_N \rightarrow$

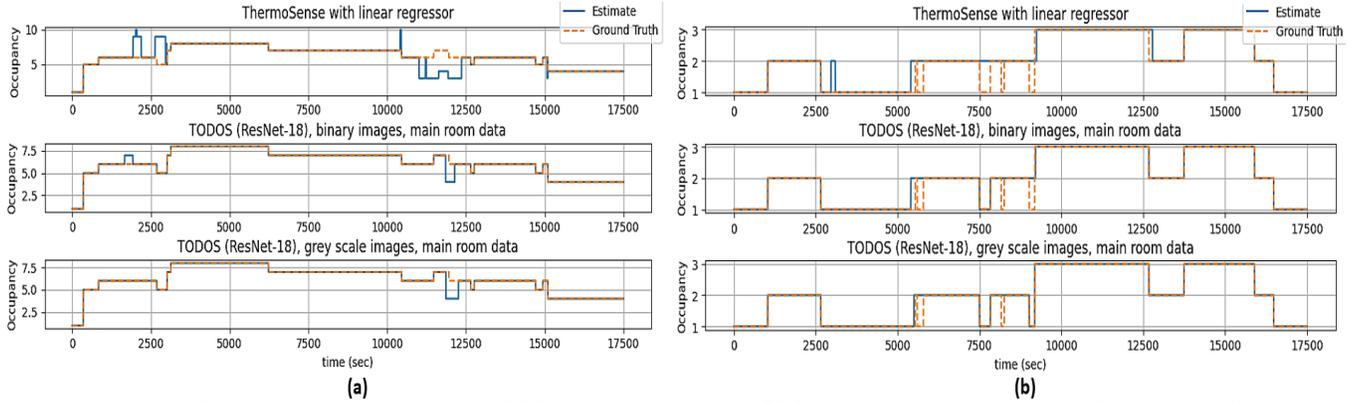


Figure 3: All sensors time-series data in (a) high occupancy, and (b) low occupancy scenario, in the main test bed.

$\mathbb{R}^D$  representing the thermal input images, and  $y_n \rightarrow \mathbb{R}$  representing the occupancy values. For hyperparameter optimization, we use a 5-fold cross-validation with grid search to estimate the generalization performance. Table 1 shows the validated hyper-parameters, the values tested by grid search, and the values that minimize the validation error. Once the optimal hyperparameter values for the neural network are obtained, we proceed to find the optimal  $\alpha$  value for the EWMA filter through another grid search. We found that a value of  $\alpha$  of 0.15 works best. Note that the best way to perform this would be to jointly optimize the cross-validation error for both the neural network and the EWMA. However, the optimization libraries available do not easily allow us to optimize both the neural network and the EWMA together as we would like. We have left this issue for future work.

## 4 PERFORMANCE EVALUATION

In this section, we proceed to experimentally evaluate TODOS performance. We considered two cases of TODOS estimates, trained with binary and greyscale images in the training set as well as concatenated thermal images with different room shapes and sizes. We also compare TODOS performance with two different closely related state-of-the-art thermal occupancy sensing schemes [6, 35], using different classifiers used in those schemes. For a thorough comparison, we tested TODOS’s performance under another standard architecture of VGG-16 [33], as well.

### 4.1 Experimental Setup

The majority of the experiments were performed in a university research laboratory of approximately  $51 m^2$  area with a rectangular shape, having 4 cubicles on each long side of the rectangle, and a meeting table in the middle with 6 chairs around it. There are also file cabinets near the entrance of the lab. Nine (9) thermal occupancy sensor motes were deployed to cover the entire lab. We considered two scenarios of low occupancy (LO, up to 3 occupants) and high occupancy (HO, up to 8 occupants), using a single sensor occupancy covering the area underneath i.e.  $2.5m \times 2.5m$ , and using all the sensors throughout the laboratory (i.e. 9) covering the entire area studied. Table 2 shows the number of experimental days with ground truth collected via direct observations of real occupancy for each scenario. During the experimental days, the room started being occupied at around 10:00 a.m., and the last occupants left at around 5:00 p.m.

Table 2: Total experimental days in the main test bed, and associated data points collected for each occupancy scenario.

	Low Occup. (LO)	High Occup. (HO)
Single Sensor	5 days $\approx$ 6,300 pts	3 days $\approx$ 3,780 pts
All Sensors	15 days $\approx$ 18,900 pts	4 days $\approx$ 5,040 pts

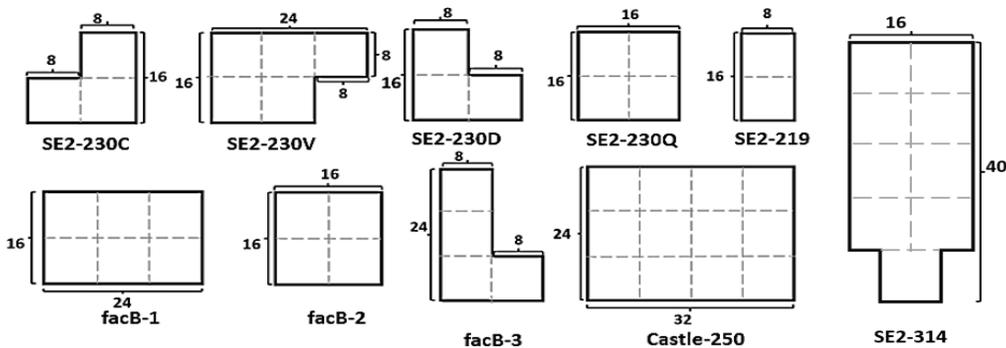
The thermal sensors were deployed on the ceiling covering the entire lab. To avoid missed detection due to blind spots, we deployed the thermal nodes sensing coverage areas very close to each other. Therefore, some sensor overlapping is inevitable when an occupant is located at the edge of two adjacent sensing areas. In this case, it may be counted twice, one by each thermal node or not counted at all, if the number of active pixels in the thermal images is not enough to trigger a count. The sensor nodes perform the background update and background subtraction operations described in § 3.2 and transmit the final thermal image to the gateway node using the low-power 802.15.4 radio. We collected occupancy sensing data every 20 seconds, and we got around 34,000 data points (i.e. thermal images) in total for our main test bed. Note that this was done to get statistically significant results. For HVAC building control, the occupancy estimate should be  $\sim 5$  to 15 minutes, to match the actuation interval commonly used in buildings.

### 4.2 Exploratory Analysis

Fig. 3 shows two examples of time-series occupancy in two days, for our main test bed in high occupancy (on the left) and low occupancy (on the right) scenarios. The top figures on each sub-figure show the occupancy estimate for the ThermoSense system using Linear Regression, the middle figures show TODOS estimates trained with binary images, and the bottom figures show TODOS estimates with greyscale images. The data is also concatenated in a single thermal image covering the entire lab as explained in § 3.3. As we see in Fig. 3, TODOS corrects for most of the occupancy estimate errors made by ThermoSense. Examples of this could be seen at  $[\sim 1800 \text{ sec}, \sim 2000 \text{ sec}]$ , and  $[\sim 11100 \text{ sec}, \sim 12000 \text{ sec}]$  in Fig. 3a, and at  $[\sim 2600 \text{ sec}, \sim 2700 \text{ sec}]$  and  $[\sim 7500 \text{ sec}, \sim 7700 \text{ sec}]$  in Fig. 3b. This is because the middle and bottom figures are obtained through data augmentation and thermal image concatenation to address the issue of sensor overlapping when an occupant is located at the edge of two adjacent sensors’ coverage areas. In both scenarios, TODOS with greyscale images provides a slightly better performance than

**Table 3: Performance evaluation in terms of occupancy estimation accuracy for different occupancy schemes under the scenarios of low (LO) and high (HO) occupancy and for two cases of using a single sensor and all sensors.**

Classifier/Regressor	Low Occupancy	Low Occupancy	High Occupancy	High Occupancy
	Single Sensor	All Sensors	Single Sensor	All Sensors
ThermoSense with Linear Reg [6]	0.99	0.93	0.95	0.91
ThermoSense with KNN [6]	0.99	0.93	0.96	0.92
Support Vector Regression (SVR) [35]	1.00	0.94	0.96	0.93
Naive Bayes [35]	0.99	0.91	0.94	0.91
Multi-layer Perceptron (MLP) [35]	1.00	0.94	0.96	0.94
TODOS w/ VGG-16, binary images	1.00	0.97	0.98	0.95
TODOS w/ VGG-16, greyscale images	1.00	0.98	0.99	0.98
TODOS w/ ResNet-18, binary images	1.00	0.98	0.98	0.95
TODOS w/ ResNet-18, greyscale images	1.00	0.99	0.99	0.99

**Figure 4: Room shapes and sensor arrangements for all the rooms tested.**

when we train it with binary images, as the greyscale image dataset is more robust against the unseen cases with its wider range of values for each thermal pixel.

### 4.3 Accuracy

Table 3 presents the classification accuracy as the main performance metric result, evaluated for each scheme tested in both low and high occupancy scenarios for single and all sensor cases. We evaluated two classifiers used in ThermoSense (i.e. linear regression and KNN), as well as Support Vector Regression, the Naive Bayes and Multi-Layer Perceptron artificial neural network (MLP) examined in [35] as the state-of-the-art techniques for comparison. In the case of support vector machines, we tried an extended version; support vector regression (SVR) with a non-linear kernel to be more flexible and robust having floating values in the output instead of integers [32]. The output will then be rounded to the nearest integer. The MLP classifier trains the model via back-propagation and the loss function is assumed to be cross-entropy in order to be comparable with the proposed classifier in this paper. We considered the same input features from the ThermoSense work (i.e. total number of active pixels, number of connected components, and size of the largest component) for all the state-of-the-art classifiers. It is noted that all results for these classifiers are obtained with no image concatenations. Also, we tested TODOS with VGG-16 architecture presented in [33] along with ResNet-18. We evaluate TODOS using these two architectures with both binary and greyscale images. All results regarding TODOS are achieved through thermal image

concatenation and data augmentation, except when using a single sensor (no concatenation).

Based on Table 3, the accuracy values for the low occupancy scenario (in both single and all sensor cases) are generally higher than the ones for the high occupancy scenario. That is due to the limited number of occupants in low occupancy scenarios that may lead to experiencing much fewer unseen cases in the training set. Considering the single sensor case alone, our proposed TODOS scheme is the only approach that delivers an accuracy of 100% in both low and high occupancy scenarios. The reason would be the data augmentation used by TODOS makes the final classifier more robust against unexpected shapes of active pixels and their combinations coming from the thermal image data. In addition, we process the input images after concatenating them based on the sensor placement in the room (automatically done based on sensor position recorded at deployment time). Furthermore, there is a more significant improvement when we use TODOS trained with greyscale images than when we train it with binary images. This way, we can correct most of the errors, especially in high occupancy scenarios for the case of all sensors (i.e. 3% and 4% improvements in VGG-16 and ResNet-18, respectively).

### 4.4 Generalization to other Rooms

In this section, we discuss and evaluate the generalization power of our TODOS model to other rooms that have no labeled training data. Different rooms have different sizes and shapes. This means that the concatenated thermal images used as input for our models are

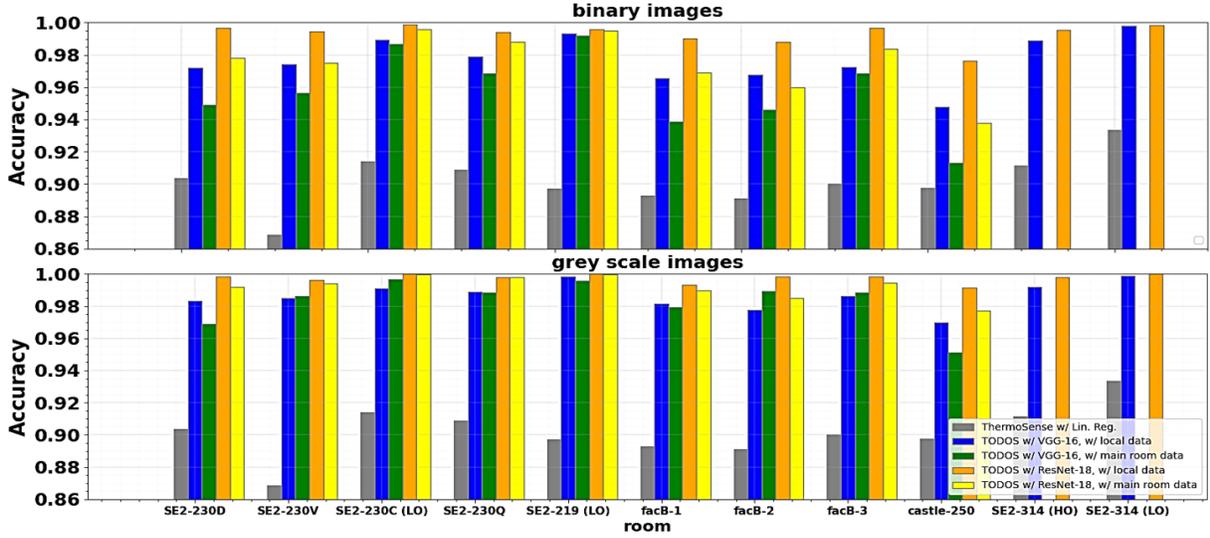


Figure 5: Accuracy results for all rooms for binary and greyscale images for VGG-16 and ResNet-18 models under two different conditions of: trained with local data, and with the main room’s data.

Table 4: Additional rooms to test the generalization of TODOS, including the area, # of data points, and # of sensors

Room	Area	# of Data Points	# of Sensors
SE2-230D	~25 m <sup>2</sup>	4 days ≈ 3,600 pts	3
SE2-230V	~25 m <sup>2</sup>	4 days ≈ 3,600 pts	5
SE2-230C	~20 m <sup>2</sup>	5 days ≈ 4,500 pts	3
SE2-230Q	~25 m <sup>2</sup>	4 days ≈ 3,600 pts	4
SE2-219	~10 m <sup>2</sup>	3 days ≈ 2,700 pts	2
facB-1	~55 m <sup>2</sup>	5 days ≈ 4,500 pts	6
facB-2	~36 m <sup>2</sup>	4 days ≈ 3,600 pts	4
facB-3	~36 m <sup>2</sup>	5 days ≈ 4,500 pts	4
castle-250	~90 m <sup>2</sup>	1 day ≈ 1,260 pts	12

different for each room/zone, which implies that we need different training data to train the models for each room. The fundamental question that we would like to answer is whether we can use labeled data obtained in a room/zone, and even a different building, and modify it to use it for training in a different room/zone.

For this, we launched an experimental campaign, deploying occupancy sensors in 10 rooms in 3 different buildings, all of different shapes and sizes. Table 4 shows the parameters of the rooms tested. Perhaps more importantly, all the rooms in the different buildings have different zone air diffusers, which could change the thermal signature detected by the sensor from building to building. Fig. 4 shows the shapes and sizes of the different rooms tested. Each of the inner squares in the rooms represents the area covered by a single sensor. For each of these rooms, we collected ground truth occupancy data, such that we could train the models with data from the specific room. We call this case, the *local data* case. In addition, we use data from the SE2-314 room (our main test bed). This room not only has the ground truth data for the total occupancy but also the occupancy sensed by each sensor in the room. Using the data from this room, we “manufacture” a training set with a subset of the data when the target room is smaller (i.e. we only take a small subset of the occupancy data representing a smaller room), or we

duplicate data when the target room is larger (i.e. we increase the data size to fit the larger room). We call this second method *main room data*. All the data used (i.e. local data and main room data) is augmented with the techniques discussed in § 3.6.

A special case happens when the target room has an irregular shape, like an L shape. In this case, the training set is of the form  $\{\vec{x}'_N, y_n\}_{n=1}^N$  with  $\vec{x}'_N \rightarrow \mathbb{R}^D$  representing the thermal input images, and  $y_n \rightarrow \mathbb{R}$  representing the occupancy values.  $\vec{x}'_N = g(\vec{x}_N) = A_{D \times D} \vec{x}_N$  with  $A$  being an indicator vector and  $\vec{x}_N$  being the training data from the main room. This vector simply allows us to zero out occupancy inputs that do not exist in the room due to its irregular shape.

Fig. 5 shows the accuracy results for all rooms for binary and greyscale images for VGG-16 and ResNet-18 models under two different conditions of (1) trained with local data, and (2) with the main room data. We also include the results of ThermoSense trained with local data for a baseline. In general, we see that grayscale images produce better results than binary images. In addition, using the local data produces more accurate results when using the ResNet-18 models. However, the results are only marginally more accurate than when using main room data (1% more error in the worst case). This discovery is *very* significant. It means that with a single curated training set, we are able to produce good enough training sets to train models in rooms of different shapes and sizes obtaining very similar performance than when using ground truth data in situ. This increases the applicability of our technique since we could potentially pre-train any model before deployment by only knowing the size and shape of the room in question.

#### 4.5 Training Data Size

Due to the non-linear nature of TODOS architecture, it is expected to have high variance in the occupancy predictions, while at the same time being able to capture relevant relations between features and target outputs. In general, this comes at the expense of requiring

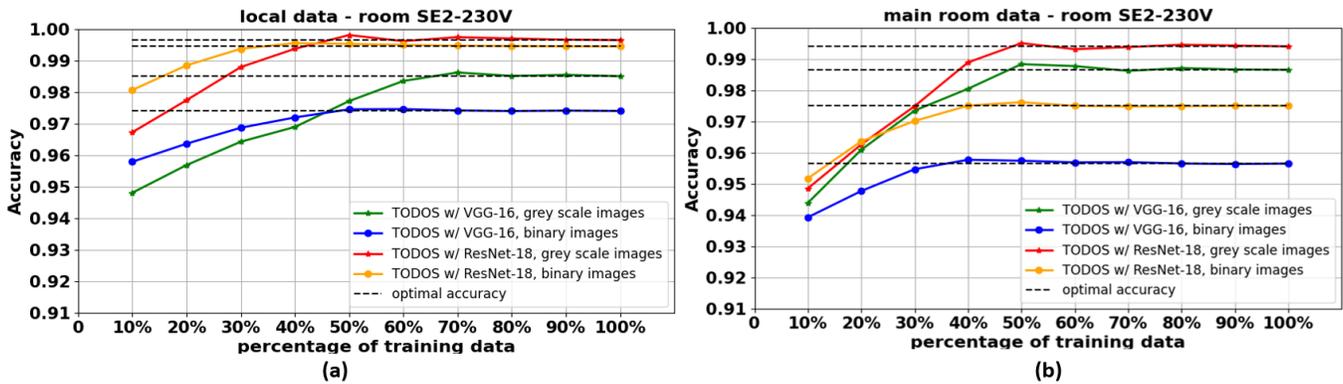


Figure 6: Accuracy as a function of the training data for room SE2-230V trained with (a) local data, and (b) main room data.

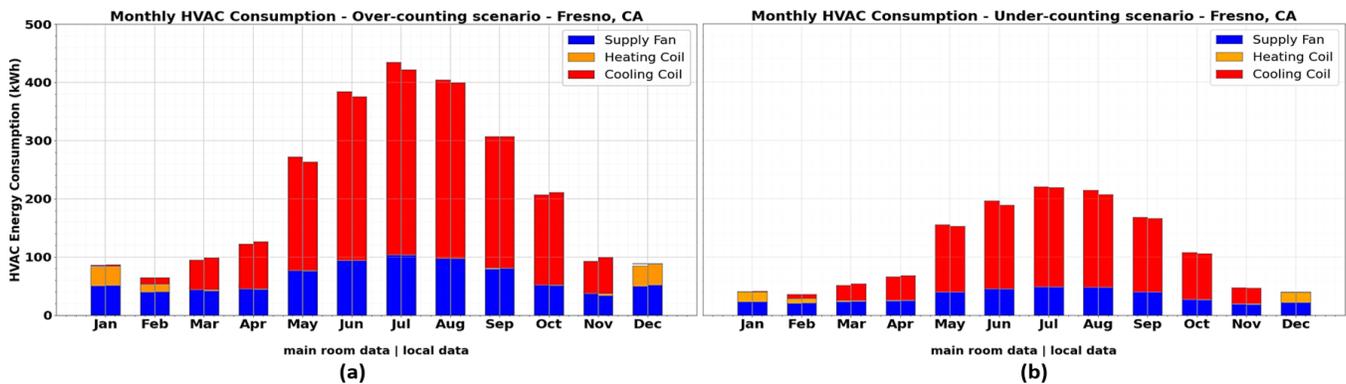


Figure 7: Monthly energy consumption for (a) over-counting, and (b) under-counting scenarios.

more training data. Thus, a good model is one that requires as little training data as possible to achieve acceptable accuracy.

Fig. 6 shows how the performance of TODOS as a function of total training data in case of being trained with the local data for a specific room (i.e. SE2-230V), and the data from our main room (i.e. SE2-314). As Fig. 6 shows, in general, TODOS with greyscale images delivers a higher accuracy at the expense of requiring more training data to converge to stable accuracy. Also, the proposed CNN architecture in TODOS (i.e. ResNet-18) provides higher accuracy with less amount of training data compared to VGG-16 architecture. This is due to the skip connections through the residual blocks in ResNet addressing the vanishing gradient issue that makes ResNet be faster and more accurate than VGG.

## 5 ENERGY AND QUALITY OF COMFORT

We analyze the energy and quality of comfort impact on HVAC building control using an occupancy-based controller. The analysis is based on an accurate occupancy estimate. The energy analysis was simulated based on: (1) the time-series occupancy information for two data schemes of TODOS trained with local data, and with the main room data, and also (2) the amount of over-/under-heating/ventilation as one important input factor. We divided the analysis into over-counting and under-counting scenarios to study the effect of false positive/negative occupancy counts on energy consumption and temperature effectiveness (as a quality of comfort

metric). We provided the input occupancy to the Energy+ simulator using the Blended Markov chain (BMC) occupancy model presented in [15]. The input model to the Energy+ includes 3786 and 3773 over-counting samples and 2252 and 2637 under-counting samples for TODOS trained with local data and main room data, respectively, with a total number of samples of 13088. Based on our Energy+ simulation results, we discuss the energy usage and quality of comfort in the following sections.

### 5.1 Energy Consumption

The impact of both data schemes (i.e. TODOS trained with local data, and with main room data) on energy usage is investigated in this section. The building's HVAC system includes a single duct terminal reheat composed of Variable Air Volume (VAV) boxes, and an Air Handler Unit (AHU). The AHU has heating and cooling coils together with a fan, that can change the air's temperature. The VAV boxes take this pre-conditioned air from the main duct and control the airflow for each zone. The power consumption sources include the supply fan, heating coils, and cooling coils. The HVAC control method is an Energy+ built-in rule-based control method based on occupied/unoccupied zone information. In the working time (i.e. 07:00 am - 06:00 pm), the heating and cooling setpoints are 21.1°C and 23.9°C, and in the non-working time (i.e. 6:00 pm - 07:00 am), they are 12.8°C and 40°C, respectively.

The monthly energy consumption for two data schemes in two scenarios of over-/under-counting is shown in Fig. 7. The Energy+ controller operates with occupancy information provided by each

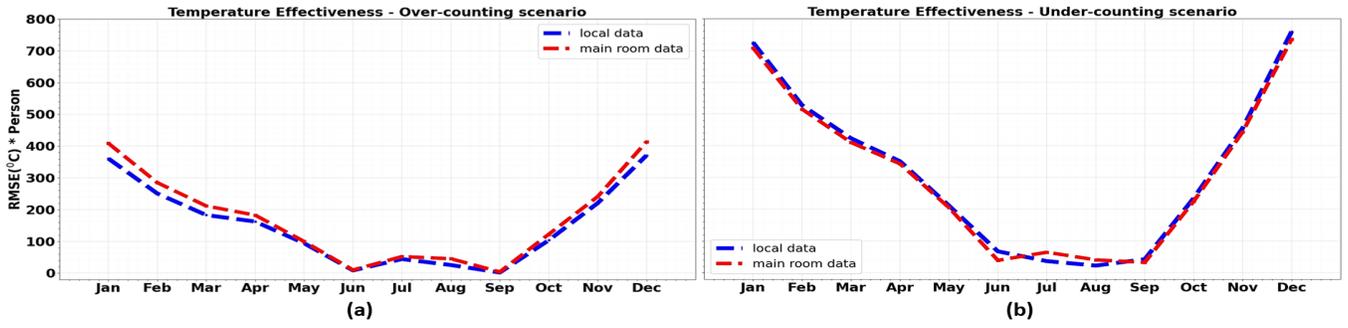


Figure 8: Monthly temperature RMSE for over-counted occupants for (a) over-counting, and (b) under-counting scenarios.

occupancy data scheme, with the over/under-counting values to the ground truth for each scheme. In the over-counting scenario, the energy consumption resulting from both data schemes are almost similar (see Fig. 7a), since those schemes have very closed values of over-counting. In this scenario, the HVAC controller will tend to consume more energy by trying to condition zones that may be empty, but TODOS occupancy estimation informs that they are occupied. In terms of under-counting, both schemes provide a very similar energy use over the entire year. In the under-counting scenario, the HVAC system will tend to float the temperature in zones that it believes to be empty, even though they are occupied. In this case, the HVAC controller consumes less energy, however, that would be at the expense of degrading the quality of comfort.

## 5.2 Quality of Comfort

In this section, we analyze the impact of both data schemes on the building’s temperature effectiveness. Temperature effectiveness is defined as the ideal temperature that should be provided to the occupants for quality of comfort. To be ASHRAE [12] compliant, the setpoint temperatures must fulfill the Predictive Mean Vote (PMV) condition:  $-0.5 \leq PMV \leq 0.5$ , where PMV is calculated by Fanger’s equation [17]. PMV predicts the mean thermal sensation vote on a standard scale for a large group of people. ASHRAE developed the thermal comfort index by using coding -3 for cold, -2 for cool, -1 for slightly cool, 0 for neutral, +1 for slightly warm, +2 for warm, and +3 for hot. PMV has been adopted by the ISO 7730 standard and it recommends maintaining PMV at level 0 with a tolerance of 0.5 as the best thermal comfort. Fanger’s PMV depends on temperature, humidity, air velocity, occupants’ clothing, and activity. We get the best temperature when the PMV is 0 (see the PMV equation above). Then, we compare the ideal temperature with the temperature under two different data schemes. For this analysis, we examine the root mean square error (RMSE) of the zone temperature difference per person between these two values.

The product of room temperature RMSE and the number of occupants for two data schemes under the over- and under-counting scenarios is shown in Fig. 8. In the over-counting scenario, both data schemes result in a better quality of comfort (i.e. smaller RMSE deviation) compared with the under-counting case. However, this higher quality of comfort comes with a high price tag, since this is obtained by over-conditioning the spaces and using a lot of energy as seen in § 5.1. In the under-counting scenario, TODOS produces a similar quality of comfort for both data schemes (even with comparable energy consumption as seen before). The energy savings

in the previous section in the under-counting scenario, come at a significant cost in the quality of comfort, as the HVAC controller will save energy by not conditioning certain zones that are in reality occupied. Therefore, that trades off the energy consumption for lower occupant quality of service.

## 6 DISCUSSION

The most important discovery was how well the models generalized by using the main room data, even in completely different rooms, different buildings, with even different air diffusers, with only a very small degradation in performance compared to using local data. We believe this works mostly because of two things. First, the type of thermal images is simpler than more complex and high-resolution images available elsewhere (e.g. web images), which makes it simpler for a well-trained neural network to recognize occupants and get very accurate counts. Second, the geometric transformations used on the augmented dataset allow compensation for different occupant body sizes, different ceiling heights, and different sensing deployment strategies, among other factors, even if they do not appear in the original labeled data. This is an important point, mostly because the transformations applied are mostly “safe”, i.e. they do not change the occupancy count, allowing to pay the price of data collection and labeling once and replicate it elsewhere.

However, there may be limits to the above. While we have tested in different buildings with different zone air diffusers and supply vents, there are many types that we did not test. It may be possible that a specific air diffuser type could change the thermal signature detected by the sensor from building to building. We believe we could compensate for these cases using a color transformation that we did not try in our data augmentation work. We have left this more detailed analysis for future work.

Another thing to discuss is that when applying data augmentation techniques for different room shapes and sizes, larger errors may be introduced due to over/under counting between adjacent training patches. For example, if we have augmented training data that consists of  $16 \times 16$ -pixel images concatenated together to form a new image of  $32 \times 16$  pixels (a new rectangular room from two squared patches), some of the image relations will be broken. While an occupant between two sensors within each square may be corrected counted, this will not be the case for an occupant between the two different squares. So a slight degradation in performance should be expected. This can be seen in the results of the Castle

Building, which showed the largest degradation in performance between the local data and the main room data (close to 1%). However, this degradation should be acceptable since it does not significantly impact the energy use and quality of comfort as shown in § 5.

## 7 CONCLUSION

In this paper, we present TODOS, a novel system for estimating occupancy in intelligent buildings. TODOS uses a low-cost, low-power thermal sensor array along with a passive infrared sensor. We introduce a novel data processing pipeline that allows us to automatically extract features from the thermal images using an artificial neural network. Through an extensive experimental evaluation, we show that TODOS provides occupancy detection accuracy of 98% to 100% under different scenarios. In addition, it solves the issue of occupancy over/under-counting produced by overlapping sensing areas when using multiple thermal sensors in large rooms. This is done by treating the entire area as a single input thermal image instead of partitioning the area into multiple thermal images individually processed. Finally, TODOS introduces a data augmentation technique that allows the generation of training data for rooms of different sizes and shapes, without requiring specific training data from each room. Using these data, TODOS can train specifically designed neural networks optimized for any room size and shape, and achieve almost the same level of occupancy detection accuracy in rooms where experimental labeled training data is available, making it a viable solution that generalizes to the different rooms in large buildings.

## 8 ACKNOWLEDGEMENTS

We would like to thank anonymous reviewers for their constructive comments and helpful suggestions. This material is based upon work partially supported by the Center of Information Technology in the Interest of Society (CITRIS).

## REFERENCES

- [1] [n. d.]. . <https://doi.org/10.1057/rlp.2010.16>.
- [2] [n. d.]. . <https://doi.org/10.1108/F-08-2019-0093>.
- [3] Stefan Achleitner, Tao Liu, Ankur U. Kamthe, and Alberto E. Cerpa. 2013. Short Term Solar Energy Forecasting Using Wireless Sensor Networks. In *Proceedings of the Second UC Solar Research Symposium*.
- [4] F. Adib, Z. Kabelac, and D. Katabi. 2015. Multi-person localization via RF body reflections. In *In Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation*.
- [5] Zhiyu An, Xianzhong Ding, Arya Rathee, and Wan Du. 2023. CLUE: Safe Model-Based RL HVAC Control Using Epistemic Uncertainty Estimation. In *ACM BuildSys*.
- [6] Alex Beltran, Varick L. Erickson, and Alberto E. Cerpa. 2013. ThermoSense: Occupancy Thermal Based Sensing for HVAC Control. In *Proceedings of the Fifth ACM Workshop on Embedded Sensing Systems for Energy-Efficient Buildings*.
- [7] N Cao, J Ting, S Sen, and A Raychowdhury. 2018. Smart sensing for HVAC Control: collaborative intelligence in optical and IR cameras. In *IEEE Trans. Ind. Electron.*, Vol. 12. IEEE, 443–455.
- [8] Veena Chidurala and Xinrong Li. 2021. Occupancy Estimation Using Thermal Imaging Sensors and Machine Learning Algorithms. *IEEE Sensors Journal* (2021).
- [9] Mozammel Chowdhury, Junbin Gao, and Rafiqul Islam. 2017. Robust human detection and localization in security applications. *Concurrency and Computation: Practice and Experience* (2017).
- [10] G. Conte and et al. 2014. a first approach using ibeacon for an energy efficient occupancy detection system. In *1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 11–19.
- [11] A. Corna and et al. 2015. Occupancy detection via ibeacon on android devices for smart building management. In *2015 Design, Automation & Test in Europe Conference & Exhibition*. 629–632.
- [12] Richard J. De Dear and Gail S. Braiger. 2002. Thermal comfort in naturally ventilated buildings: revisions to ASHRAE Standard 55. *Energy and Buildings* (2002).
- [13] Xianzhong Ding, Wan Du, and Alberto Cerpa. 2019. Octopus: Deep reinforcement learning for holistic smart building control. In *ACM BuildSys*.
- [14] Xianzhong Ding, Wan Du, and Alberto E Cerpa. 2020. Mb2c: Model-based deep reinforcement learning for multi-zone building control. In *ACM BuildSys*.
- [15] Varick Erickson, Miguel A. Carreira-Perpinan, and Alberto E. Cerpa. 2011. OBSERVE: Occupancy-based system for efficient reduction of HVAC energy. In *ACM/IEEE IPSN*.
- [16] European Commission. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance).
- [17] Povel O. Fanger. 1970. Thermal comfort. Analysis and applications in environmental engineering. *Copenhagen: Danish Technical Press*. (1970).
- [18] G. Fierro and et al. 2012. Zone-level occupancy counting with existing infrastructure. In *Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. ACM, 205–206.
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *CoRR* (2015). arXiv:1512.03385
- [20] M. Jin, N. Bekiaris-Liberis, K. Weekly, NC.J. Spanos, and A.M. Bayen. 2002. Occupancy detection via environmental sensing. *IEEE Trans. Autom. Sci. Eng.* (2002).
- [21] Ankur Kamthe, Lun Jiang, Matthew Dudys, and Alberto Cerpa. 2009. SCOPES: Smart Cameras Object Position Estimation System. In *EWSN 2009: Wireless Sensor Networks*. European Conference on Wireless Sensor Networks, 279–295.
- [22] Neil E Klepeis et al. 2001. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *Journal of Exposure Science and Environmental Epidemiology* 11, 3 (2001), 231–252.
- [23] Kiran Maharana, Surajit Mondal, and Bhushankumar Nemade. 2022. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings* (2022). ICIEA-2022.
- [24] Ubaid Mehmood, I. Moser, Prem Prakash Jayaraman, and Abhik Banerjee. 2019. Occupancy Estimation using WiFi: A Case Study for Counting Passengers on Busses. In *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*.
- [25] Mostafa Mirshekari, Jonathon Fagert, Shijia Pan, Pei Zhang, and Hae Young Noh. 2021. Occupant Localization in Obstructive Indoor Environments Using Footstep-Induced Floor Vibrations. In *Dynamics of Civil Structures*. Springer.
- [26] Nicola Moretti, J. D. Blanco Cadena, A. Mannino, T. Poli, and F. Re Cecconi. 2021. Maintenance service optimization in smart buildings through ultrasonic sensors network. *Intelligent Buildings International* (2021).
- [27] Abdallah Naser, Ahmad Lotfi, and Junpei Zhong. 2021. Adaptive Thermal Sensor Array Placement for Human Segmentation and Occupancy Estimation. *Buildings and environments* 21, 2 (2021), 1993–2002.
- [28] Abdallah Naser, Ahmad Lotfi, and Junpei Zhong. 2021. Occupancy determination based on time series of CO2 concentration, temperature and relative humidity. *Buildings and environments* 21, 2 (2021), 1993–2002.
- [29] Shijia Pan, Amelie Bonde, Jie Jing, Lin Zhang, Pei Zhang, and Hae Young Noh. 2014. Boes: building occupancy estimation system using sparse ambient vibration monitoring. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2014*. International Society for Optics and Photonics.
- [30] Shijia Pan, Mostafa Mirshekari, Pei Zhang, and Hae Young Noh. 2016. Occupant traffic estimation through structural vibration sensing. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2016*.
- [31] Hamid Rajabi, Zhizhang Hu, Xianzhong Ding, Shijia Pan, Wan Du, and Alberto Cerpa. 2022. MODES: Multi-Sensor Occupancy Data-Driven Estimation System for Smart Buildings. In *ACM e-Energy*.
- [32] Bernhard Schölkopf, Alex J. Smola, Robert C. Williamson, and Peter L. Bartlett. 2000. New Support Vector Algorithms. *Neural Computation* (2000).
- [33] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs.CV]
- [34] Robert Tomastik, Yiqing Lin, and Andrzej Banaszuk. 2008. Video-based estimation of building occupancy during emergency egress. In *IEEE ACC*.
- [35] Ash Tyndall, Rachel Cardell-Oliver, and Adrian Keating. 2016. Occupancy Estimation Using a Low-Pixel Count Thermal Imager. *IEEE Sensors Journal* (2016).
- [36] Junqi Wang, Jingjing Huang, Zhuangbo Feng, Shi-Jie Cao, and Fariborz Haghighat. 2021. Occupant-density-detection based energy efficient ventilation system: Prevention of infection transmission. *Energy and Buildings* (2021).
- [37] Z.Y. Wu, W. Lv, and K. Yu. 2016/10. A Framework of Intelligent Evacuation Guidance System for Large Building. In *ICCAHE 2016*.
- [38] Hengyang Zhao et al. 2018. Thermal-Sensor-Based Occupancy Detection for Smart Buildings Using Machine-Learning Methods. *ACM Transactions on Design Automation of Electronic Systems* (2018).
- [39] Han Zoua, Yuxun Zhoua, Hao Jiangc, Szu-Cheng Chiend, Lihua Xie, and Costas J. Spanos. 2018. WinLight: A WiFi-based occupancy-driven lighting control system for smart building. *Energy and Buildings* 158 (2018), 924–938.
- [40] M.S. Zuraimi et al. 2018. Predicting occupancy counts using physical and statistical Co2-based modeling methodologies. *Buildings and environments* (2018).