Enhancing the detection performance of a vision-based occupancy detector for buildings

Author 1

Paige Wenbin Tien, PhD Researcher

- Department of Architecture and Built Environment, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom

ORCID Number: 0000-0003-0123-248X

Author 2

- Shuangyu Wei, PhD Researcher
- Department of Architecture and Built Environment, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom

Author 3

- Tin Wai Chow, MSc
- Department of Architecture and Built Environment, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom

Author 4

- Jo Darkwa, Professor, PhD, MSc.
- Department of Architecture and Built Environment, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom

Author 5

- Christopher Wood, Associate Professor, PhD, MSc
- Department of Architecture and Built Environment, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom

Author 6

- John Kaiser Calautit, Associate Professor, PhD, MSc
- Department of Architecture and Built Environment, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom

Full contact details of corresponding author.

Email: paige.tien@gmail.com, paige.tien@nottingham.ac.uk

Abstract

Occupant behaviour is one of the key parameters that significantly impact the operation of heating, ventilation, and air-conditioning (HVAC) systems and the energy performance of buildings. The detailed occupancy information can improve HVAC operation and utilisation of building spaces. Strategies such as vision-based occupancy detection and recognition have recently garnered much interest. This study investigates the performance of a vision-based deep learning detection technique for enhancing building system operations and energy performances. The model used was the Faster RCNN with Inception V2. Two occupancy detection model configurations were developed, tested and evaluated. Both models were analysed based on the application of the detector within a selected case study building, along with the evaluation based on the different evaluation metrics. Results suggest that the occupancy detector (Model 1) provided an overall accuracy of 95.23% and an F1 score of 0.9756, while the occupancy activity detector (Model 2) provided an accuracy of 89.37% with an F1 score of 0.8298. Building Energy Simulation (BES) was used to evaluate and compare the impact of such an approach on the indoor occupancy heat gains. The study highlighted the potential of the detection approaches, but further development is necessary, including optimisation of the model, full integration with HVAC controls and further model training and field testing.

Keywords chosen from ICE Publishing list

Built environment; Energy; Artificial intelligence; Artificial neural network; Buildings

Nomenclature

API	Application Programming Interface
BES	Building Energy Simulation
COCO dataset	Common Objects in Context Dataset
CO2	Carbon Dioxide
CNN	Convolutional Neural Network
DLIP	Deep Learning Influenced Profile
HVAC	Heating, Ventilation and air-conditioning

IoU	Intersection over Union
mAP	Mean Average Precision
PC	Personal Computer
R-CNN	Region-based Convolutional Neural Network
RFID	Radio Frequency Identification
SSD	Single-shot Detector

1. Introduction and Literature Review

Occupancy behaviour and patterns within building spaces have been identified as significant factors impacting building energy efficiency (Delzendeh et al., 2017). Recent studies have investigated occupancy behaviour in buildings and developed demand-driven solutions to improve building system operations (Paone and Baacher, 2018). To obtain occupancy data, various technologies were employed, including infrared (Yun and Lee, 2014), Wi-Fi (Simma et al., 2019) and Radio Frequency Identification (RFID) (Li et al., 2011). These solutions provide information about a building space, such as occupancy count and location, however, there are several limitations, such as the requirement of multiple sensors distributed across the room and limitations in terms of recognising occupancy behaviour (activities, interaction with equipment or appliances) and the determination of the location of the occupants within the space (Dongre et al. 2019). Furthermore, indirect methods such as environmental-based sensors were used (Yun and Won, 2012), to monitor the changes within the space when occupants are present. Effectively, the data collected is employed to develop demand-driven solutions for more effective system controls (Kathirgamanathan et al., 2021), energy optimisation (Salimi and Hammad, 2020), and also building energy management (Jin et al., 2018). Many of these solutions are based on artificial intelligence and machine learning models that have advantages in terms of adaptability and application to different types of buildings (Amasyali and El-Gohary, 2018). Studies suggest that further enhancement of such strategies should include achieving a multi-objective system that enables building energy and comfort management (Shaikh et al. 2018).

Other solutions, such as data-driven and forecasting-based methods, can also be used to optimise the operation of building energy systems but are dependent on historical data or patterns (Marinakis, 2020). However, the diversity in occupancy among different spaces and varying occupancy activities in buildings may present challenges for such solutions. Furthermore, a potential time delay can occur between the prediction and the provision of the actual building requirements. This indicates the need to develop solutions such as demand-driven controls that can adapt to varying occupancy patterns in real-time and optimise HVAC operations.

Furthermore, the cooling/heating design setpoint temperature assigned to building spaces is usually based on the indoor space's purpose/function. For instance, the CIBSE Guide (CIBSE, 2015) suggests operative temperatures for spaces such as offices, libraries and restaurants at 21 - 25°C in the UK. Additionally, conventional building HVAC systems are typically operated based on fixed or predefined scheduled profiles. The impact of different occupancy patterns and activities are typically not considered, resulting in over or under conditioned building spaces.

To resolve such limitations, recently there has been an increase in research employing computer vision and deep learning-based approaches that enable real-time detection and recognition in buildings to reduce building energy demands (Tien et al., 2022 and Wei et al., 2021). The studies employed a computer vision approach to detect and predict the internal heat gains in office buildings based on the detected occupancy and activities. It was highlighted that the occupancy behaviour directly impacted the energy consumed in buildings (Tien et al., 2021a) and also, indirectly (Tien et al. 2021b), via internal heat gains from the use of electrical equipment or appliances such as computers and monitors (Wei et al., 2021). Furthermore, such vision-based approaches can also be used to detect the operation of windows in buildings (Tien et al., 2021c) and indoor fires (Pincott et al., 2022). The predicted information can be used to adjust the control and operation of the HVAC to reduce the energy demand and enhance thermal comfort (Wang et al., 2022). In addition, it can generate realistic occupancy profiles for building energy models, potentially reducing the performance gap. These are initial studies that introduced the framework and approach, with no in-depth investigation of the model configuration and its impact on the performance of the detection model, in particular, the data curation, labelling and training employed. Furthermore, the impact of selecting a suitable response category; between the ability to detect and recognise the number of occupants, position, and activities performed must be investigated. Finally, most of the studies focus on small office spaces and the performance of the detector must be evaluated when applied in larger spaces and number of occupants.

2. Method

To enable accurate and real-time detection of occupancy levels and activities within building spaces to assist the operations of building energy systems, the present work will employ the method introduced in (Tien et al., 2021c). Two different occupancy detector configurations were developed, tested, and analysed. Detailed performance comparisons are provided through the real-time application of the detectors within a selected case study building, and the use of different evaluation metrics. Furthermore, a comparison between the actual observation (ground truth) and the generated occupancy profiles also called here deep learning influenced profiles (DLIP), was carried out. Building Energy Simulation (BES) was used to show if the occupancy heat gains could be represented more accurately using the two occupancy detector configurations as compared to ground truth.

2.1 Case Study Building

A postgraduate study space on the first floor of the Paton House Building at the University Park Campus, University of Nottingham, UK was selected to assist in testing the developed real-time occupancy detectors. The Paton House Building is a typical Victorian-style house (Qu et al., 2021) which was repurposed by the University as teaching and office spaces. The climate in the case study area can be classified as a temperate oceanic (Kottek et al., 2006). The location and images of the Paton House Building are shown in Figures 1a and 1b. The building is naturally ventilated and integrated with a central heating system. The U-values of external walls, external floor, roof, doors, and windows are 1.42, 0.95, 1.46, 2.33, and 5.20 W/m^{2}K, respectively. The test room has a floor area of 36.62 m^2 and a floor-to-ceiling height of 3.52 m, and there are six sliding sash windows that can be opened at the bottom for ventilation. Figure 1c presents the floor plan of the first floor of the building along with the configuration of the room shown in Figure 1d. To enable the capture of the whole test room, cameras with a resolution of 1080p and a wide 90-degree field of view was fixed in the corner of the room and close to the ceiling. It should be noted that this case study building is not intended to evaluate the building itself or its facilities but rather for testing the detection methods in a small-size classroom with occupants performing activities common in this type of space.

Figure 1: Paton House at University Park Campus, University of Nottingham, UK. (a) Location map of the building, (b) Outlook of the building, (c) Building floor plan, (d) Room configuration, (e, f) Experimental test room setup.

As shown in Figures 1e and 1f, two cameras were placed at two corners of the room with Detection Camera A, and Detection Camera B. Figure 2 presents the field of view from both cameras. For this study, the detection performance evaluation was only carried out using Camera B. Furthermore, the room has a capacity limit due to COVID-19 restrictions during the test period. For the experimental test, there were 8 participants. The detection performance analysis was based on the detection and recognition of each participant, as shown in Figure 2. It should be noted that in practice, images/videos of occupants are not saved during the real-time detection and are only shown here for demonstration purposes.

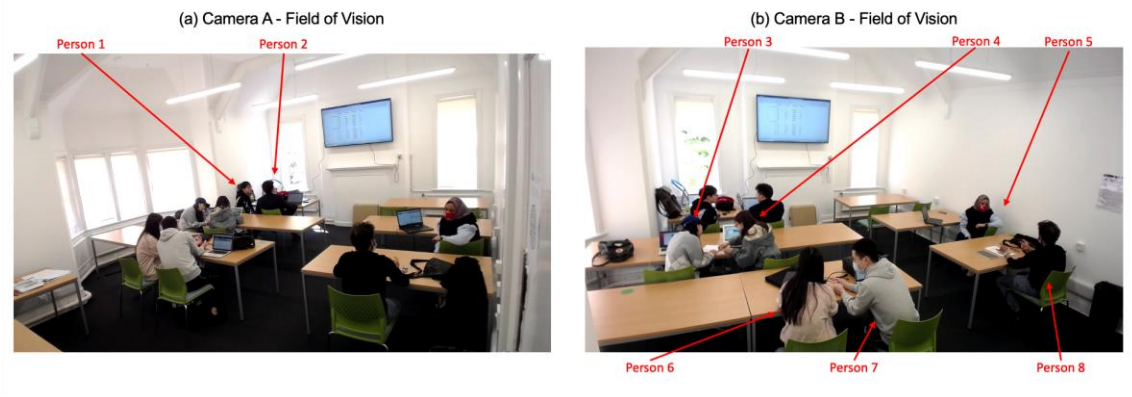


Figure 2: Field of view from Camera A & B with the identification of occupants 'People 1 – 8' for the purposes of detection performance analysis.

2.2 Development of the Vision-based Detector Using a Deep Learning Method

Recently, many studies have focused on employing data and demand-driven solutions to enhance HVAC operation and performance (Kallio et al., 2021). The application of vision-based techniques for detection and recognition tasks using a camera device has many advantages, but at the same this has limitations. Using vision-based systems in indoor spaces presents several challenges, including obstacles blocking the view of the desired detection area or objects. Internal environmental conditions, including lighting and glare, could impact detection and recognition performance.

The framework introduced in (Tien et al., 2022) highlights the potential of using deep learning techniques based on a classification-based algorithm to develop computer vision-based detectors. It showed the potential of using detected occupancy information to assist HVAC system controls. The present study will build on previous knowledge and technique to establish two occupancy detection model configurations and evaluate their capabilities.

Following the model development process 2 in (Tien et al., 2022), images of occupants were collected to form the datasets described in Table 1. Model 1 is configured to detect the number of people in the space. Whereas Model 2 is configured to detect and recognise common occupancy activities performed by the occupants. This includes ‘sitting’, ‘standing’ and ‘walking’.

Table 1: Number of images and labels per category for Models 1 and 2.

Category	Number of Images			Number of Labels		
	Training	Testing	Total	Training	Testing	Total
Model 1: People Counting						
People	40	10	50	168	45	213
Model 2: Occupancy Activities						
Sitting	400	100	500	753	149	902
Standing	400	100	500	701	134	835
Walking	400	100	500	1000	177	1177
Total	1200	300		2454	460	

The software, LabelImg (Tzutalin, 2015) was used to label all of the images located within both datasets manually. As shown in Figure 3, labels were assigned entirely around each specific region of interest. For most images, more than one occupant appears within the image; hence multiple labels were assigned.

Figure 3: Example images from the training and testing image datasets used to train the models.

In this study, a Convolutional Neural Network (CNN) model configuration was used. To assist the development of the neural network, the TensorFlow Object Detection API was used. This framework platform provides pre-trained models through a transfer learning approach that enables the development of the vision-based occupancy detector. Existing models provided in the TensorFlow Detection Model were explored to establish the model configurations. Based on the assessment of the different models, the pre-trained model Faster R-CNN (with Inception V2) was selected. The time required for training the models would vary due to the differences in the input data and the desired detection output responses.

The trained models were deployed to a camera to provide real-time detections in the selected postgraduate study space. A scenario consisting of eight occupants present within the space was recorded. This ensured that the two detection model configurations were evaluated using the same sequence of occupancy activities and positions. It also ensured that other factors such as the indoor lighting conditions and glare did not influence the results, providing a fair comparison between the model's detection and recognition abilities. The detection and recognition responses were obtained and recorded every second, generating the DLIP.

For Model 1, count-based profiles were generated, giving the number of occupants detected over time within the building space. For Model 2, similar profiles generation process following (Tien et al., 2022) was employed; with three responses, sitting, standing and walking. The formed DLIPs would be assessed and compared with the true 'actual observation' to evaluate the overall performance of each occupancy detector.

The model's performance was assessed based on the average Intersection over Union (IoU) accuracy (Wu et al., 2020), the percentage of the time achieving correct, incorrect and no detections and the confusion matrix. Further evaluation was performed based on the common metrics of precision, recall and F1 score. Details about these evaluation metrics are detailed in (Goutte and Gaussier, 2005) and employed in similar studies (Tien et al., 2022 and Wei et al., 2022).

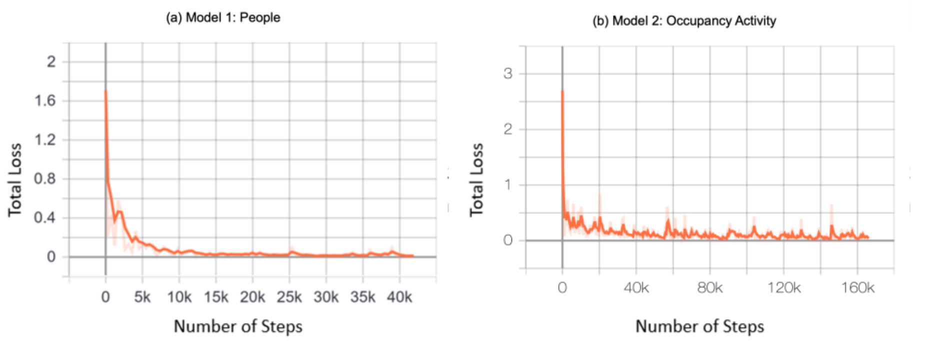
3. Results and Discussion

The following section presents the results of the training and testing of the two different occupancy detectors. It presents the analysis of the performance of the detector during an field experiment conducted within the selected building space. A further evaluation was by comparing the DLIP profiles generated with the ground truth results or the actual observation profiles.

3.1 Training

A summary of the training results is given in Table 2. Since Model 1 had a smaller image dataset and only one response assigned, it led to a shorter training duration and fewer training steps than Model 2. As observed in Table 2, the total loss versus the training steps plot indicates the complexity of Model 2 compared to Model 1. Greater fluctuations were seen during the model training. Effectively, based on the loss convergence, both models were trained and should be able to carry out the detection tasks.

Table 2: Training results for the two occupancy detectors.

Training Conditions and Results	**Model 1: People Detection Model**	**Model 2: Occupancy Activity Model**
Model Used	Faster RCNN with Inception V2	
Total Steps	41,901	166,128
Training Duration	2 hours, 54 minutes	10 hours, 29 minutes, 52 seconds
Average Loss	0.07607	0.13436
Minimum Loss	0.003567	0.005654
Total loss versus the number of training steps		

To confirm the completion of the training of the models, initial detection was performed using the test images from the dataset. Results are presented in Table 3 with the confusion matrix and the common classification metrics. The confusion matrix presents the ability of the two classification models (Model 1 and 2) based on their performance of a set of testing data whereby the true values are known. Model 1 was designed to only recognise one type of response (people) via a binary classification problem, while Model 2 has a total of 3 detection responses with the addition of none/other classification to assist the analysis of the model performance. For both models, true positive results were achieved when the classifier correctly recognises the person present in the building space and true negative when it correctly recognises no people in the space. The confusion matrix also presents the amount of false positive and false negative results achieved, referring to the number of detections that were incorrectly detected. Based on the confusion matrix, the walking activity achieved a higher accuracy (92.66%) compared to the other activities of sitting (87.92%) and standing (82.84%). However, the standing and walking activities may have similar occupancy body form and shape, which could present difficulties in identifying the true activity, it led to the occurrence of walking being incorrectly identified as standing (11.19%). Overall, the results showed the detectors' potential as effective occupancy detectors. To further evaluate the trained models in terms of their ability to classify occupancy and activities, common evaluation metrics, including accuracy, precision, recall and F1 score were used (Sokolova et al. 2006).

Table 3: Confusion matrix and model performance results based on the evaluation of the model using the test image dataset.

Confusion Matrix

(a) Model 1

Predicted Class \ True Class	Person	No Person/ Other
Person	82.76%	1.72%
No Person/ Other	13.79%	1.72%

(b) Model 2

Predicted Class \ True Class	Sitting	Standing	Walking	None/ Other
Sitting	87.92%	5.22%	0.56%	1.34%
Standing	4.03%	82.84%	4.52%	0.00%
Walking	3.36%	11.19%	92.66%	0.00%
None/ Other	3.36%	0.75%	2.26%	-

Class	Accuracy	Precision	Recall	F1 Score
Model 1: People				
People	84.48%	0.9796	0.8571	0.9143
Model 2: Occupancy Activities				
Sitting	94.04%	0.925	0.8911	0.9077
Standing	91.43%	0.9064	0.8284	0.8657
Walking	92.70%	0.8643	0.9266	0.9047
Average	92.72%	0.8986	0.8820	0.8927

Figure 4 presents snapshots of the detection and recognition during the experimental test using the two different occupancy detection models. Figures 4a, b, c shows the results achieved from the application of Model 1, and Figures d to i for Model 2. For the majority of time, Model 1 enabled the detection and recognition of most occupants within the building space. Whereas Model 2, had some no/false incorrect detections in identifying the occupancy activities. Many of these instances occurred directly for the occupants furthest away from the camera and/or obstructed by objects in the room or by other people. Further analysis will be given to identify the benefits and limitations of each of the detection model configurations.

Figure 4: Snapshots of occupancy detection and recognition during various key stages of the experimental test using the different detectors.

Figure 5 shows that Model 1 achieves an average detection IoU of 98.85% for all the occupants. Despite Occupancy 6 within the direct view and angle of the camera, a slightly lower IoU (93.60%) was achieved. This may have resulted from the participant facing opposite the camera in most instances. Future works should take this into account when creating the training dataset. However, overall the results indicate the ability of the vision-based detection approach to enable real-time identification of the number of occupants present in a building space.

During the experimental test, the activity of sitting was performed by all occupants. For this activity, consistent IoU was achieved, with an average IoU accuracy of 92.80%. Only some of the occupants performed the standing and walking activities. Hence, further evaluation of other activities must be carried out in future works. The results showed IoU accuracies of 85.25% and 71.25% were achieved for standing and walking activities. Such a lower IoU accuracy was due to the difficulty in detecting and recognising these two types of activities with similar occupancy body form and shape. The results in Figure 5B also suggest that the IoU accuracy was not highly impacted by the different occupants in the space and their positions in relation to the camera, indicating the detection camera was positioned at a suitable place within the room to capture the activities of most of the occupants.

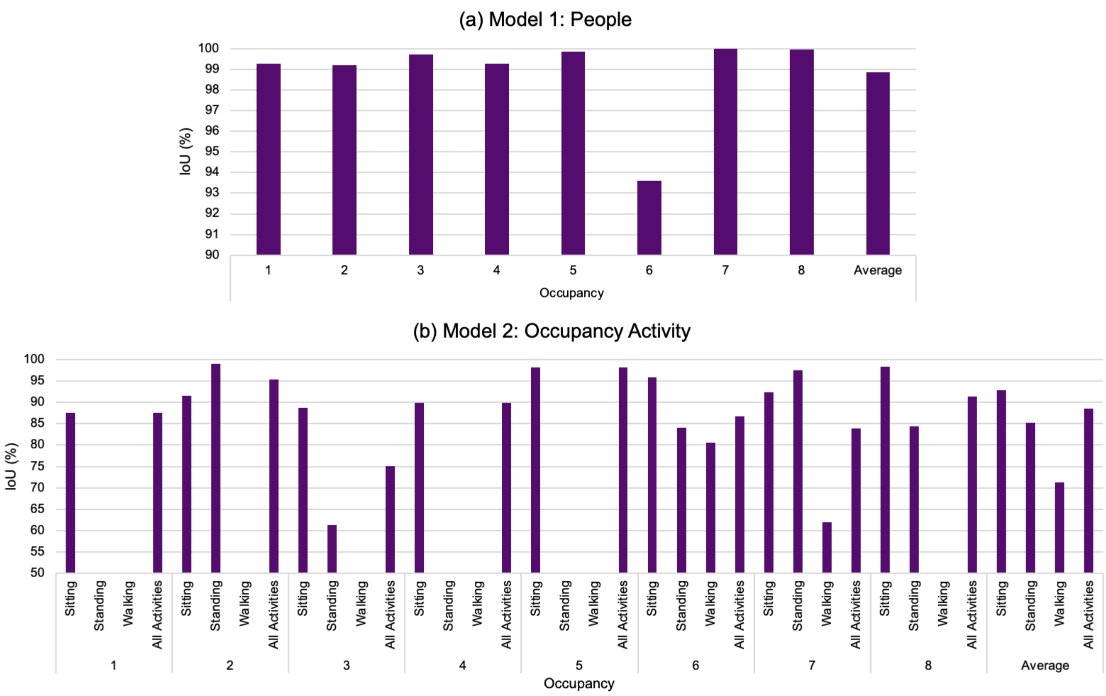


Figure 5: Average IoU (%) of the occupants during the experimental test using Models 1 and 2.

A detailed summary of the results is presented in Table 4. The results suggest that achieving correct/incorrect and no detections would have been influenced by the model performance on recognising each occupant within the space. For Model 1, up to 100% correct detection could be achieved along with minimal incorrect detections. No/ missed detections also occurred.

For Model 2, which detects occupancy activities, the results suggest that the detection performance was varied across each occupant. It should be noted that not all occupants performed all types of activities. Overall, for all three activities, the percentage of correct detections was the highest, with an average of 74.13%, compared to incorrect detections at 1.25% and no/missed detections at 24.63%. The highest no/missed detections were observed for occupant 1, with a no/missed detection rate of 64.12%. This may be due to occupant 1 being one of the furthest from the camera.

Table 4: Detection performance in terms of the percentage of time achieving correct, incorrect, and no detections.

Percentage of Time Achieving:				
Model 1: People				
Occupant (People Detection)		**Correct Detections**	**Incorrect Detections**	**No/ Missed Detections**
1		89.70%	0.00%	10.30%
2		76.41%	0.00%	23.59%
3		99.67%	0.00%	0.33%
4		97.34%	0.00%	2.66%
5		99.67%	0.00%	0.33%
6		99.67%	0.00%	0.33%
7		99.34%	0.33%	0.33%
8		100.00%	0.00%	0.00%
Average		95.22%	0.04%	4.73%
Model 2: Occupancy Activities				
Occupant	**Activity**	**Correct Detections**	**Incorrect Detections**	**No/ Missed Detections**
1	Sitting	35.55%	0.33%	64.12%
	Standing	N/A	N/A	N/A
	Walking	N/A	N/A	N/A
	All Activities	35.55%	0.33%	64.12%
2	Sitting	56.04%	0.00%	43.96%
	Standing	33.33%	33.33%	33.33%
	Walking	N/A	N/A	N/A
	All Activities	55.81%	0.33%	43.85%
3	Sitting	67.59%	1.03%	31.38%
	Standing	36.36%	63.64%	0.00%
	Walking	N/A	N/A	N/A
	All Activities	66.45%	3.32%	30.23%
4	Sitting	65.12%	0.00%	34.88%
	Standing	N/A	N/A	N/A
	Walking	N/A	N/A	N/A
	All Activities	65.12%	0.00%	34.88%
5	Sitting	98.01%	0.00%	1.99%
	Standing	N/A	N/A	N/A
	Walking	N/A	N/A	N/A

	All Activities	98.01%	0.00%	1.99%
6	Sitting	92.25%	0.00%	7.75%
	Standing	50.00%	41.67%	0.00%
	Walking	100.00%	0.00%	0.00%
	All Activities	91.03%	1.66%	7.31%
7	Sitting	81.88%	3.83%	14.29%
	Standing	92.31%	7.69%	0.00%
	Walking	100.00%	0.00%	0.00%
	All Activities	82.39%	3.99%	13.62%
8	Sitting	98.98%	0.00%	0.00%
	Standing	85.71%	14.29%	0.00%
	Walking	N/A	N/A	N/A
	All Activities	98.67%	0.33%	1.00%
Average	Sitting	74.43%	0.65%	24.80%
	Standing	59.54%	32.12%	6.67%
	Walking	100.00%	0.00%	0.00%
	All Activities	74.13%	1.25%	24.63%

To further evaluate the performance of the detectors during the experimental tests, Figure 6 and Figure 7 present the results in the form of the confusion matrix. For model 1, the results verify the results presented in Table 4 with the lowest true positives values of 76.41%, and the highest number of false positives of up to 23.59% was for the detection of Occupant 2. In comparison to the detection of the other occupants, more consistent results were achieved, giving minimal false negatives with no false positives. Overall, an average of 95.23% were achieved for true positives in correctly detecting people within the space.

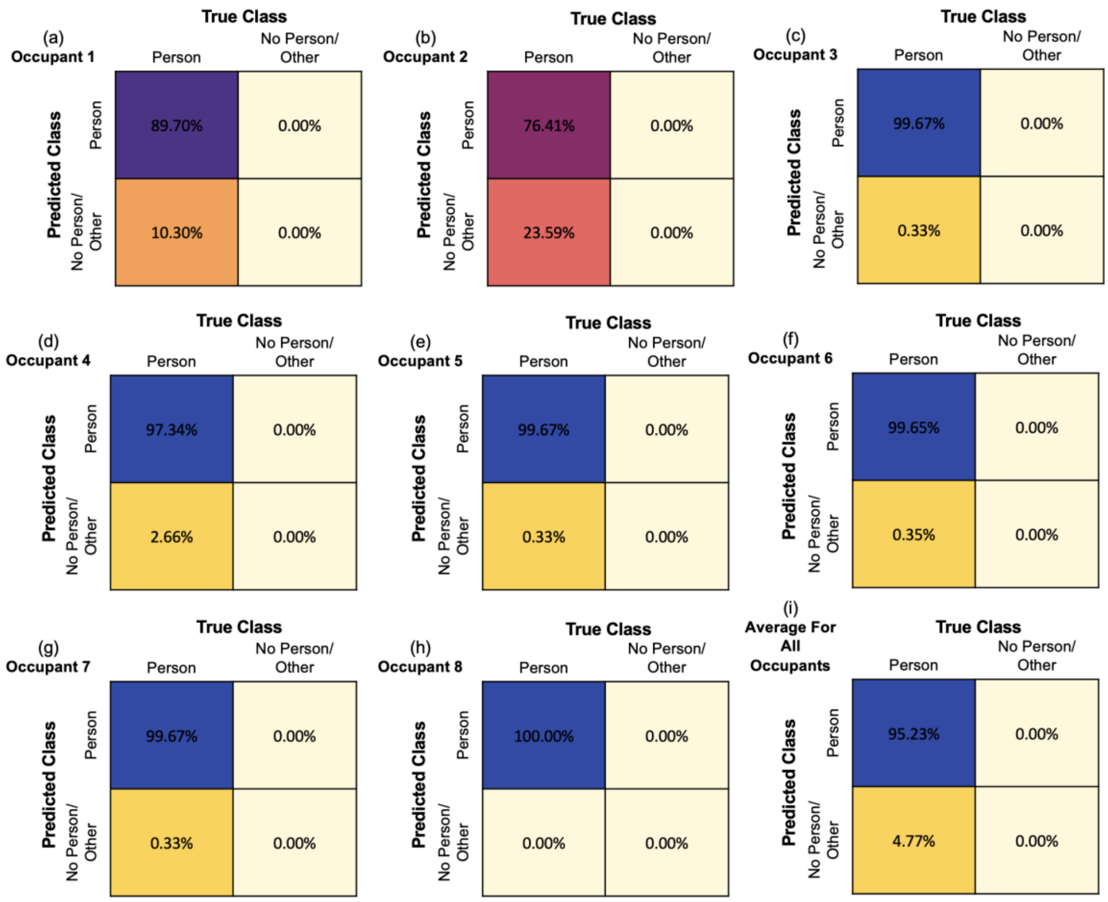


Figure 6: Detection performance results for Model 1 (people detector) in the form of a confusion matrix.

As presented in Figure 7, the results suggest Model 2 can adequately identify each of the different activities performed by the occupants. The results indicate that the walking activity achieved the most true positives, with a value of up to 100%. Secondly, it is followed by the sitting activity. This achieved up to an average of 74.18%. The confusion matrix for each occupant suggests that the lower percentage achieved for this activity was due to the occasion of no prediction when this activity was performed. Furthermore, the standing activity was sometimes predicted as sitting and/or no detection of such activity, giving the worst performance compared to the other responses. The overall performance shown in Figure 7i was used to calculate the common evaluation metrics, including the accuracy, precision, recall, and the associated F1 scores given in Tables 5 and 6 for both models.

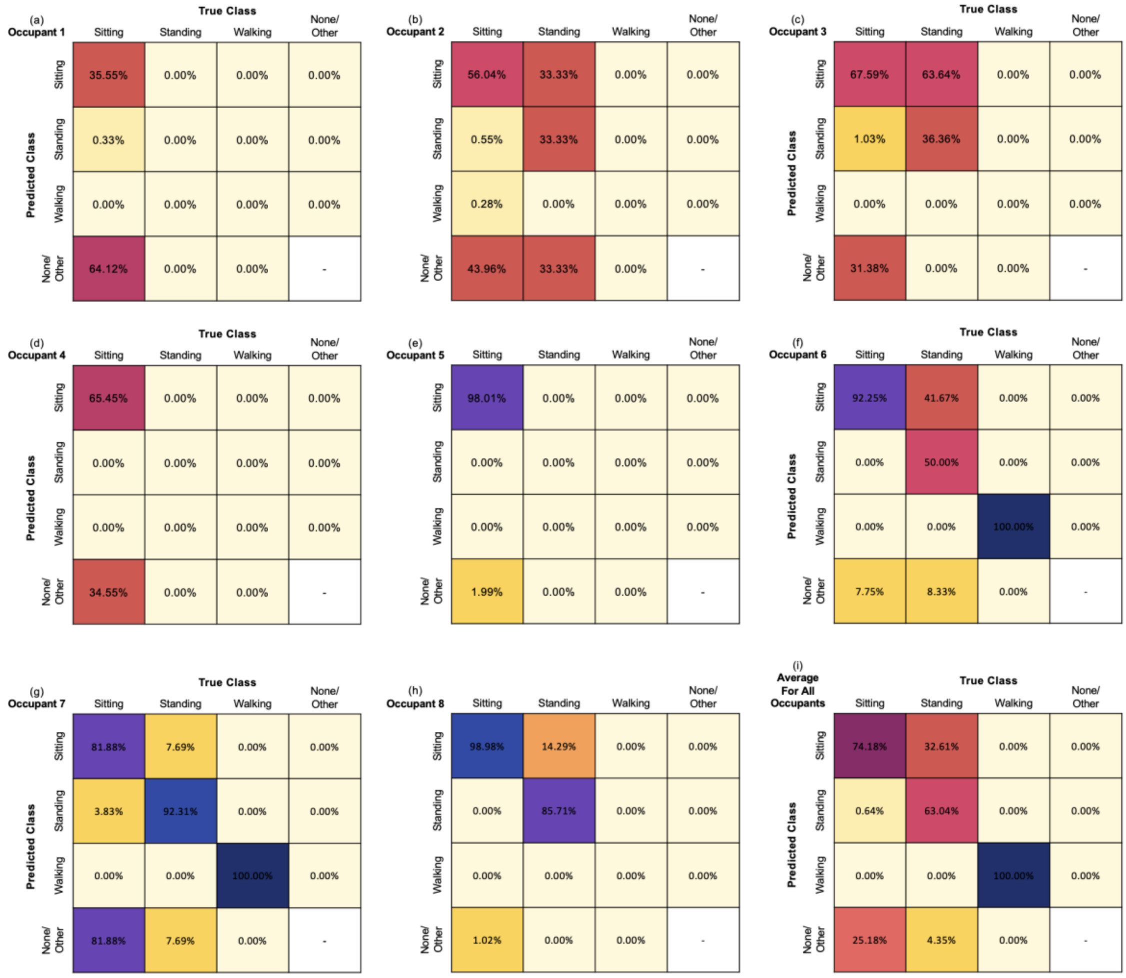


Figure 7: Detection performance results for Model 2 (occupancy activity detector) in the form of a confusion matrix.

The evaluation metrics results are shown in Table 5. Model 1 provided an overall accuracy of 95.23% and an F1 score of 0.9756. Model 2 provides an accuracy of 89.37% with an F1 score of 0.8298. Since multiple responses were selected for this model, further development is required to ensure a consistent level of detection accuracy could be achieved across the different occupancy activities. Furthermore, since both models were only tested on a selected experimental test, further analysis is required to evaluate whether both models can effectively assist the operations of building HVAC systems and enhance the building energy performances through further testing on different indoor spaces and variation in variation occupancy conditions. For example, Model 1 may be effective in predicting the CO2 concentration levels based on the occupancy count, while Model 2 would be more suitable for evaluating the heat gains from occupants or predicting the activity rate for thermal comfort calculations in real-time.

Table 5: Detection performance results based on common classification evaluation metrics from the application of Models 1 and 2.

Occupant	Class	Accuracy	Precision	Recall	F1 Score
Model 1: People					
1	Person	89.70%	1.000	0.8970	0.9457
2	Person	76.41%	1.000	0.7641	0.8663
3	Person	99.67%	1.000	0.9967	0.9983
4	Person	97.34%	1.000	0.9734	0.9865
5	Person	99.67%	1.000	0.9967	0.9983
6	Person	99.65%	1.000	0.9965	0.9982
7	Person	99.67%	1.000	0.9967	0.9983
8	Person	100.00%	1.000	1.000	1.000
Average	Person	95.23%	1.000	0.9523	0.9756
Model 2: Occupancy Activity					
1	Sitting	35.55%	1.0000	0.3556	0.5245
	Standing	N/A	N/A	N/A	N/A
	Walking	N/A	N/A	N/A	N/A
	All Activities	35.55%	1.0000	0.3556	0.5245
2	Sitting	61.35%	0.6270	0.5604	0.5918
	Standing	66.67%	1.0000	0.3334	0.5001
	Walking	N/A	N/A	N/A	N/A
	All Activities	64.01%	0.8135	0.4469	0.5460
3	Sitting	51.98%	0.5150	0.6759	0.5846
	Standing	67.67%	0.9725	0.3636	0.5293
	Walking	N/A	N/A	N/A	N/A
	All Activities	59.83%	0.7438	0.5198	0.5570
4	Sitting	64.65%	1.0000	0.6545	0.7912
	Standing	N/A	N/A	N/A	N/A
	Walking	N/A	N/A	N/A	N/A
	All Activities	64.65%	1.0000	0.6545	0.7912
5	Sitting	98.01%	1.0000	0.9801	0.9800
	Standing	N/A	N/A	N/A	N/A
	Walking	N/A	N/A	N/A	N/A
	All Activities	98.01%	1.0000	0.9801	0.9800
6	Sitting	83.53%	0.6888	0.9225	0.7887
	Standing	83.33%	1.0000	0.5000	0.6667
	Walking	100.00%	1.0000	1.0000	1.0000
	All Activities	88.95%	0.8963	0.8075	0.8185
7	Sitting	91.40%	0.9141	0.8188	0.8638
	Standing	96.16%	0.9602	0.9231	0.9413
	Walking	100.00%	1.0000	1.0000	1.0000
	All Activities	95.85%	0.9581	0.9140	0.9350
8	Sitting	92.35%	0.8738	0.9898	0.9282
	Standing	92.86%	1.0000	0.8571	0.9231
	Walking	N/A	N/A	N/A	N/A
	All Activities	92.61%	0.9369	0.9235	0.9257
Average	Sitting	80.64%	0.6975	0.7418	0.7190
	Standing	87.47%	0.9899	0.6304	0.7703
	Walking	100.00%	1.0000	1.0000	1.0000
	All Activities	89.37%	0.8958	0.7907	0.8298

Based on the experimental test, the detections and recognitions were recorded in the form of the DLIPs. Figure 8 presents the generated count-based profiles for each response achieved using Model 1 and their activities in Model 2. Since the same video recording was used for testing the models, it essentially compared the number of occupants present in a building space versus the occupants performing various activities.

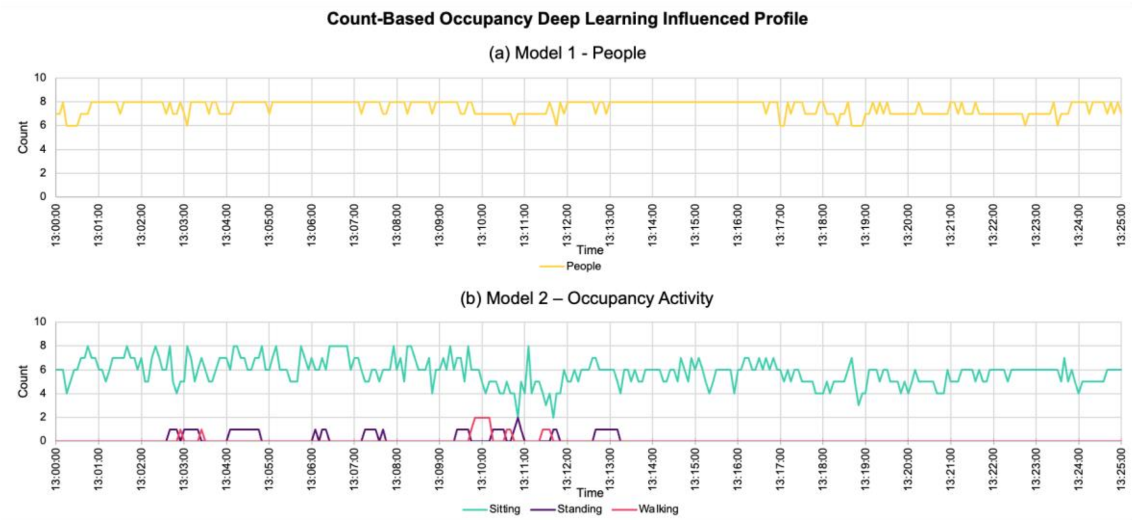


Figure 8: Formed deep learning influenced profiles (DLIPs) from the application of Models 1 and 2 during the experimental test.

The ground truth or Actual Observation profile was used to further assess the detection performance of the methods. Results given in Figure 9 suggest that the occupancy and activities profiles consistently fluctuate, indicating prediction error. Therefore, further improvements are required to enhance the detection model's accuracy, reliability, and stability.

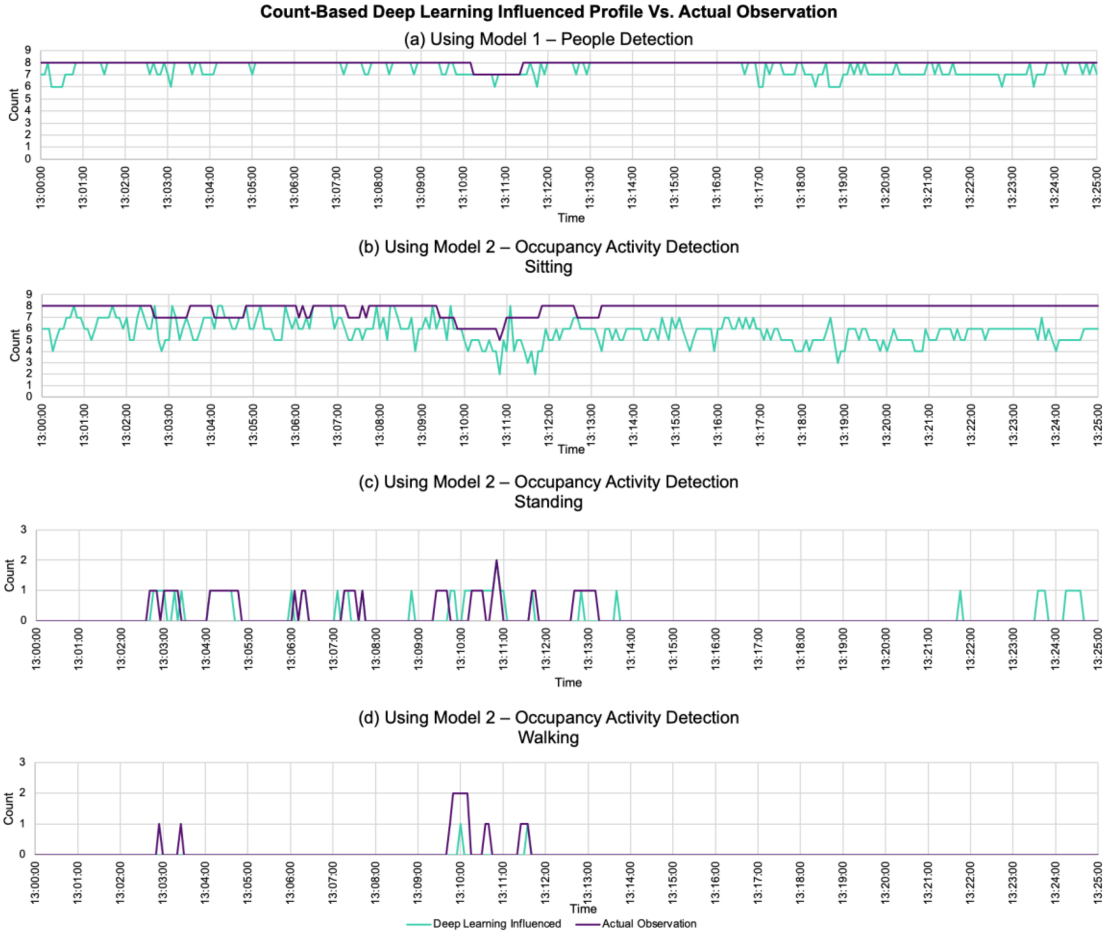


Figure 9: Comparison of the formed DLIPs with the Actual Observation Profile.

Figure 10 compares the generated DLIP with Actual Observation profiles (ground truth). The occupancy count DLIPs generated using Model 1 was used to predict the occupancy heat emissions. Heat gains profiles were generated assuming the detected occupants were sitting. Moreover, Figure 10 also presents the predicted occupancy heat emission profile based on the occupancy activity profiles in Figure 8b, generated using Model 2. This was compared with Model 1, indicating a difference in heat emissions of up to 29.75%. As compared with the Actual Observation profile, a difference of up to 5.69% for Model 1 and 25.36% for Model 2 was observed, indicating substantial errors in the detections of Model 2.

Although Model 2 was supposed to be more accurate in predicting the actual heat emissions of the occupants, Model 1 was closer to the Actual Observation (ground truth) due to the limited activities (mostly seating) performed by the occupants during the experimental test. This highlights the importance of developing an accurate and stable occupancy activity detector in order to be effective and valuable for building control systems. Furthermore, a greater impact could potentially be observed when such a detection method is implemented within larger indoor spaces with more people performing various occupancy activities.

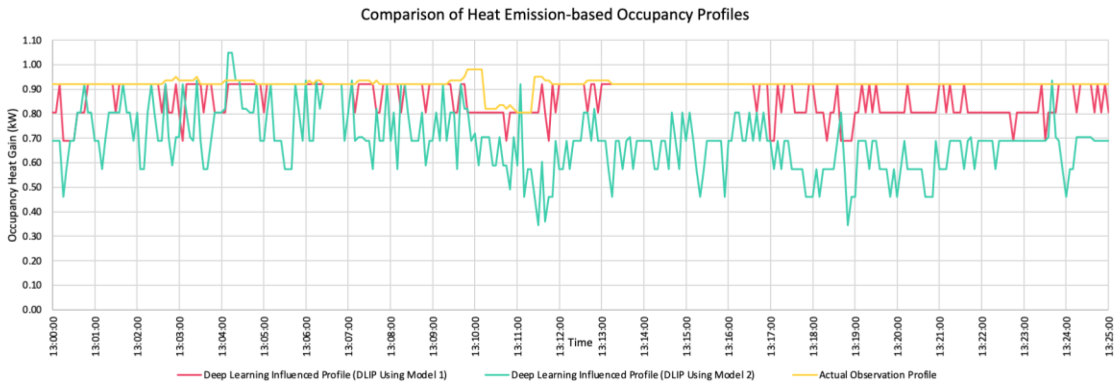


Figure 10: The generated DLIP using Models 1 and 2 plotted against predefined and the Actual Observation Profile.

4. Conclusion and Future Works

The study investigates the development of a vision-based deep learning detection technique for enhancing building system operations and energy performance. Two occupancy detection approaches based on Faster RCNN with Inception V2 model were developed, tested and evaluated. Model 1 focused on detecting the number of occupants in a building space. While Model 2 focused on detecting common occupancy activities such as ‘sitting’, ‘standing’ and ‘walking’. Similar images were used for training the model, and the same training procedure was conducted. Both models were evaluated based on an experimental test performed within a postgraduate study space at the University.

Model 1 provided an overall accuracy of 95.23% and an F1 score of 0.9756, providing good detection of the number of occupants within the indoor space. Model 2 provided a lower accuracy of 89.37%, with an F1 score of 0.8298. Since Model 2 had multiple detection tasks, further development is required to ensure a consistent level of detection accuracy could be achieved across the different occupancy activities. Although Model 2 was supposed to be more accurate in predicting the actual heat emissions of the occupants, Model 1 was closer to the Actual

Observation (ground truth) due to the limited activities (mostly seating) performed by the occupants during the experimental test. This highlights the importance of developing an accurate and stable occupancy activity detector in order to be effective and valuable for building control systems. It is envisaged that the proposed detection approach could have a greater impact when applied in a larger indoor space with more occupants and different types of activities. Hence, future works should evaluate the application of the detection approach in various types of indoor spaces with variations in the number of occupants and their activities. Further model training with larger datasets should be carried out to improve the overall detection performance. The impact of parameters such as the indoor lighting conditions and positioning of the detection camera should be evaluated.

Acknowledgements

The authors would like to acknowledge the suppsupport from the Department of Architecture and Built Environment, University of Nottingham and the PhD studentship from EPSRC, Project References: 2100822 (EP/R513283/1).

References

AMASYALI, K. & EL-GOHARY, N.M. 2018. A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews, 81*, Part 1, 1192-1205.

CIBSE Chartered Institution of Building Services Engineer. 2015. Environmental design: CIBSE Guide A. In. London: *CIBSE*.

DELZENDEH, E., WU, S., LEE, A., ZHOU, Y. 2017. The impact of occupants' behaviours on building energy analysis: A research review. *Renewable and Sustainable Energy Reviews*, 80, 1061-1071.

DONGRE, P, ROOFIGARI-ESFAHAN, N., DEJONG, M.J., SCHOOLING, J.M., VIGGANI, G.M.B. 2019. Occupant-building interaction (OBI) model for University *buildingsInternational Conference on Smart Infrastructure and Construction 2019 (ICSIC). January 2019*, 631-637.

GOUTTE, C., GAUSSIER, E. 2005. Aprobabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: Losada D.E., Fernández-Luna J.M. (eds) *Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science*, vol 3408. Springer, Berlin, Heidelberg.

JIN, X., WANG, G., SONG, Y., SUN, C. 2018. Smart building energy management based on network occupancy sensing. *Journal of International Council on Electrical Engineering*, 8:1, 30-36.

KALLIO, J., TERVONEN, J., RÄSÄNEN, P., MÄKYNEN, R., KOIVUSAARI, J., PELTOLA, J. 2021. Forecasting office indoor CO2 concentration using machine learning with a one-year dataset. *Building and Environment*, 187, 107409.

KATHIRGAMANATHAN, A., DE ROSA, M., MANGINA, E., FINN, D.P. 2021. Data-driven predictive control for unlocking building energy flexibility: A review. *Renewable and Sustainable Energy Reviews*, 135, 110120.

KOTTEK, M., GRIESER, J., BECK, C., RUDOLF, B., RUBEL, F. 2006. World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15:3, 259-263.

LI, S., Li, N., BECERIK-GERBER, B., CALIS, G. 2011. RFID-based occupancy detection solution for optimizing HVAC energy consumption. *2011 Proceedings of the 28th ISARC, Seoul, Korea*, 587-592.

MARINAKIS, V. 2020. Big data for energy management and energy-efficient buildings. *Energies* 13:7, 1555.

PAONE, A., BACHER, J. 2018. The impact of building occupant behavior on energy efficiency and methods to influence it: A review of the state of the art. *Energies*, 11:4, 953.

PINCOTT, J., TIEN, P.W., WEI, S., CALAUTIT, J.K. 2022. Development and evaluation of a vision-based transfer learning approach for indoor fire and smoke detection. *Building Services Engineering Research and Technology*, 43:3, 319-332.

QU, K., CHEN, X., WANG, Y., CALAUTIT, J., RIFFAT, S., CUI, X. 2021. Comprehensive energy, economic and thermal comfort assessments for the passive energy retrofit of historical buildings – A case study of a late nineteenth-century Victorian house renovation in the UK. *Energy*, 220, 119646.

SALIMI, S. & Hammad, A. 2020. Optimizing energy consumption and occupants comfort in open-plan offices using local control based on occupancy dynamic data. *Building and Environment*, 176, 106818.

SHAIKH, P.H., NOR, N.B.M., NALLOGOWNDEN, P., ELAMVAZUTHI, I. 2018. Intelligent multi-objective optimization for building energy and comfort management. *Journal of King Saud University-Engineering Sciences*, 30:2, 195-204.

SIMMA, K.C.J., MAMMOLI, A., BOGUS, S.M. 2019. Real-time occupancy estimation using WiFi network to optimize HVAC operation. *Procedia Computer Science*, 155, 495-502.

SOKOLOVA, M., JAPKOWICZ, N., SZPAKOWICZ, S. 2006. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In: Sattar, A., Kang, Bh. (eds) AI 2006: Advances in Artificial Intelligence. AI 2006. Lecture Notes in Computer Science(), vol 4304. Springer, Berlin, Heidelberg.

TIEN, P.W., WEI, S., CALAUTIT, J.K, DARKWA, J, WOOD, C. 2021a. A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand driven control solutions. *Energy and Buildings*, 226, 110386.

TIEN, P.W., WEI, S., CALAUTIT, J.K. 2021b. A computer vision-based occupancy and equipment usage detection approach for reducing building energy demand. *Energies*, 14, 156.

TIEN, P.W., WEI, S., CALAUTIT, J.K, DARKWA, J, WOOD, C. 2021c. A deep learning approach towards the detection and recognition of opening of windows for effective management of building ventilation heat losses and reducing space heating demand. *Renewable Energy, 177, 603-625.*

TIEN, P.W., WEI, S., CALAUTIT, J.K., DARKWA, J, WOOD C. 2022. Real-time monitoring of occupancy activities and window opening within buildings using an integrated deep learning-based approach for reducing energy demand. *Applied Energy*, 308, 118336.

TZUTALIN, 2015. LabelImg, Available: https://github.com/tzutalin/labelImg. [Accessed February 2022].

WANG, Z., CALAUTIT, J., WEI, S., TIEN. P.W., XIA, L. 2022. Real-time building heat gains prediction and optimization of HVAC setpoint: An integrated framework. *Journal of Building Engineering*, 49, 104103.

WEI, S., TIEN, P.W., CALAUTIT, J.K., WU, Y., BOUKHANOUF, R. 2021. Vision-based detection and prediction of equipment heat gains in commercial office buildings using a deep learning method. *Applied Energy*, 277, 115506.

WEI, S., TIEN, P.W., WU, Y., CALAUTIT, J.K. 2022. A coupled deep learning-based internal heat gains detection and prediction method for energy-efficient office building operation. *Journal of Building Engineering*, 47, 103778.

WU, S., LI, X., WANG, X. 2020. IoU-aware single-stage object detector for accurate localization. *Image and Vision Computing*, 97, 103911.

YUN, J., Lee, S. 2014. Human movement detection and identification using pyroelectric infrared. *Sensors*, 14, 8057-8081.

YUN, J., WON, K.H. 2012. Building environment analysis based on temperature and humidity for smart energy systems. *Sensors*, 12:10, 12458-13470.