Research paper

BUILDING SERVICES ENGINEERING RESEARCH & TECHNOLOGY

Building Serv. Eng. Res. Technol.
2021, Vol. 0(0) 1–23

DOI: 10.1177/01436244211026120
journals.sagepub.com/home/bse
SAGE

Vision-based human activity recognition for reducing building energy demand

Paige Wenbin Tien, Shuangyu Wei, John Kaiser Calautit, Jo Darkwa and Christopher Wood

Abstract

Occupancy behaviour in buildings can impact the energy performance and the operation of heating, ventilation and air-conditioning systems. To ensure building operations become optimised, it is vital to develop solutions that can monitor the utilisation of indoor spaces and provide occupants' actual thermal comfort requirements. This study presents the analysis of the application of a vision-based deep learning approach for human activity detection and recognition in buildings. A convolutional neural network was employed to enable the detection and classification of occupancy activities. The model was deployed to a camera that enabled real-time detections, giving an average detection accuracy of 98.65%. Data on the number of occupants performing each of the selected activities were collected, and deep learning–influenced profile was generated. Building energy simulation and various scenario-based cases were used to assess the impact of such an approach on the building energy demand and provide insights into how the proposed detection method can enable heating, ventilation and air-conditioning systems to respond to occupancy's dynamic changes. Results indicated that the deep learning approach could reduce the over- or under-estimation of occupancy heat gains. It is envisioned that the approach can be coupled with heating, ventilation and air-conditioning controls to adjust the setpoint based on the building space's actual requirements, which could provide more comfortable environments and minimise unnecessary building energy loads.

Practical application: Occupancy behaviour has been identified as an important issue impacting the energy demand of building and heating, ventilation and air-conditioning systems. This study proposes a vision-based deep learning approach to capture, detect and recognise in real-time the occupancy patterns and activities within an office space environment. Initial building energy simulation analysis of the application of such an approach within buildings was performed. The proposed approach is envisioned to enable heating, ventilation and air-conditioning systems to adapt and make a timely response based on occupancy's dynamic changes. The results presented here show the practicality of such an approach that could be integrated with heating, ventilation and air-conditioning systems for various building spaces and environments.

Department of Architecture and Built Environment, University of Nottingham, Nottingham, UK

Corresponding author:
Paige Wenbin Tien, Department of Architecture and Built Environment, University of Nottingham, University Park, Nottingham NG7 2RD, UK.
Email: paige.tien@gmail.com, paige.tien@nottingham.ac.uk

Keywords

Deep learning, occupancy activity detection, building energy management, heating, ventilation and air-conditioning systems, building energy performance

Received 15 February 2021; Revised 28 April 2021; Accepted 1 June 2021

Introduction and literature review

Buildings are responsible for a significant percentage of global emissions and energy consumption.[1] With the increasing urbanisation and population and improving quality of life, buildings' energy consumption is expected to keep rising.[2] A large portion of the energy consumed in buildings is due to heating, ventilation and air conditioning (HVAC).[3] Therefore, many researchers are focussing on developing solutions such as passive strategies, advanced and smart controls. Although reducing the HVAC's energy consumption will go a long way towards developing a low-carbon future, it is also important to consider the occupant's comfort when creating new solutions. An example of this is the use of intelligent and demand-driven controls for HVACs,[4] which can reduce or eliminate unnecessary energy consumption while still providing comfortable and healthy indoor spaces.[5,6]

Occupants' behaviour can significantly impact the operation of HVAC and building energy demand.[7] Heating, ventilation and air conditioning, which uses conventional control strategies or fixed setpoint schedules, could not adjust to the conditioned spaces' actual requirements. For example, an office HVAC system operating with fixed schedules will assume a certain number of occupants present during the working day. Hence, it could condition spaces that are not fully occupied or completely unoccupied leading to rooms being under- or over-conditioned and may lead to a substantial waste of resources.[5,6] This is even more important now with the COVID-19 restrictions, with many offices following staggered shifts to reduce workplace congestion.[8] Several studies[9,10] have shown that the daily occupancy rate on average can be below 60%, in particular, in single-person offices.

Employing HVAC control strategies that can coordinate real-time usage of building services to occupants' presence, such as demand-driven building controls, is necessary.[11] These solutions reduce energy usage and enhance thermal comfort by optimising the scheduling of the operation of HVAC by using the occupancy information.[5,9] For example, demand-driven control strategies can adjust the heating and cooling operation of HVACs according to the actual requirement of building spaces.[12]

However, the effective development and implementation of such control strategies require accurate and real-time data on occupants' presence in building spaces.[13] The presence of occupants can be detected and monitored using sensing technologies.[14] Many of the studies[14–16] focussed on detecting the number and location of occupants in buildings. Another factor that should be considered is the occupants' activities within the space, which can, directly and indirectly, influence the internal sensible and latent heat gains[15,16] but only limited works have been carried out so far. An occupant sitting in space will have a lower heat emission compared to the same person who is more active, that is, walking or carrying an object. The capability to detect and recognise the occupants' activities would be useful to provide information that can be used to better identify the actual heating/cooling requirements of a space.[17] Employing artificial intelligence (AI) methods can help develop accurate detection and recognitions tools.[18]

An example of an AI method employed to carry out tasks such as object classifications and visual and speech recognition with great accuracy is deep learning.[19–21] With deep learning, many new applications of computer vision methods have been introduced. Recently, several deep learning and computer vision methods have been implemented in the building sector to enhance systems and operations. For example, the study[22] developed a video-based method to detect occupants' number in a space for energy conservation applications. The results showed that the detection accuracy for occupancy measurement was 95%; however, the study did not provide detailed results regarding its impact on energy consumption. The work of Markovic et al.[23] also explored the potential of the deep learning technique for conserving the energy for the building sector but focussed on predicting the opening and closing of windows. The results showed that the evaluation accuracy was 86–89%. These studies show the capabilities of such an approach to accurately collect information about the space in real-time and then use it to make decisions and changes to the building services operation to optimise energy efficiency and comfort.

Most of the relevant previous works are mainly focussed on enhancing the model's performance, such as its accuracy, speed, etc., in detecting the number and distribution of occupants in a space. To date, only limited studies have attempted to demonstrate the usage of the information generated from such an approach to control the operation of HVAC for optimising energy efficiency and thermal comfort. In addition, only limited studies[8,15] have employed such solutions to detect and recognise the occupant's activities and predict the heat emitted to the space by the occupants (sensible and latent heat gains). Furthermore, the impact of the approach implementation on the building energy consumption should be explored.

The present work will address the previously mentioned gaps in research by developing a coupled approach based on computer vision and deep learning to detect and classify the occupancy patterns, including the occupant count, location and activities in a building space, in real time. More specifically, based on previous works,[15] we propose a technique based on the Faster regions with CNN features (R-CNN) that enables detection and recognition of occupancy patterns and usage of equipment in an office space (Figure 1). The model is trained and deployed to a standard camera, and field tests were carried out in an office space. During the field test, the proposed method's capabilities are evaluated by detecting multiple occupants performing different activities in an office space such as walking, sitting and standing.

Different evaluation metrics suggested by previous works were employed. The field tests were carried out in an actual working space in a building at the university. The acquired real-time information is used to generate occupancy profiles which can then be used to control HVAC operation and as input for building energy simulation (BES) models. In this study, the proposed approach is evaluated by simulating the case study building in BES with the generated profiles. Several scenarios were developed and simulated to further evaluate the approach. This includes comparing the approach with the use of fixed schedules.

Method

The following section presents an outline of the proposed framework approach and the methods to

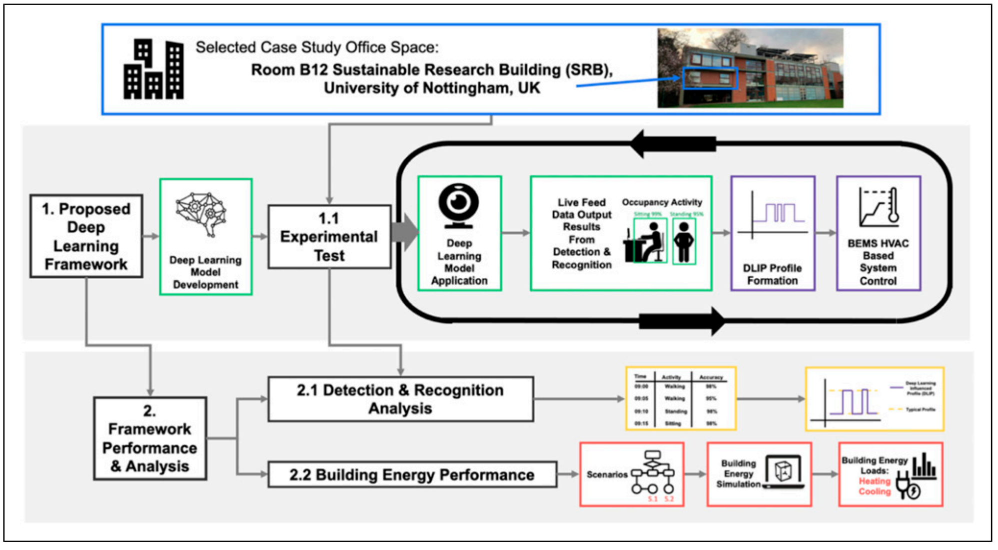


Figure 1. Overview of the research method and framework approach.

develop, assess and evaluate a vision-based deep learning occupancy activity detection and recognition approach.

Methodology overview

This present study aims to use a BES tool to analyse the impact of applying the proposed vision-based real-time occupancy detection approach on the energy demand. The first part discusses the deep learning techniques employed to develop the vision-based approach to detect and recognise occupancy activities. It consisted of the configuration and training of a convolutional neural network (CNN)-based model using an image dataset. The trained model was then deployed to an AI-powered camera and was used to perform live detections. As highlighted in Figure 1, an experimental test was carried out within a selected office space environment to evaluate the effectiveness and performance of the proposed approach. Data on the number of occupants performing different activities were collected, and a deep learning–influenced profile (DLIP) was generated.

The second part analyses the trained model's performance in terms of detection and recognition accuracy during the experimental test. Furthermore, the selected case study office building was modelled using the BES tool, and the proposed method was compared against the use of typical or fixed occupancy operation schedules. This was carried out to assess the potential influence on the building energy demand and HVAC systems' operations.

Computer vision with deep learning method

One of the most common applications of deep learning for computer vision is object detection. Object detection has been extensively developed and is popular for applications in terms of video surveillance,[24] crowd counting[25] and also anomaly detection[26] due to its unique ability to locate objects within an image or video. By incorporating image classification tasks with localisation, object detection outperforms image classification due to its ability to perform well under detections with multiple objects in the image of different types.[27,28]

The workflow consists of selecting a suitable deep learning method that enables occupancy activity recognition within an indoor office-based environment. In the present study, the CNN was employed due to its good performance in tasks involving images and videos.[29] Next, data in the form of images were gathered and pre-processed, and the model would be further configured prior to training. Subsequently, the trained model is deployed to an AI-powered camera, ready to perform detection. The following section presents further details about the workflow stages.

Datasets and pre-processing stages. To establish the vision-based deep learning occupancy detection model, input data in the form of images were collected and pre-processed to the desired format, ready for training. Common occupancy activities performed within typical office buildings were selected as the desired model detection responses. Table 1 presents the number of images collected and the number of labels assigned within both the training and testing datasets. This dataset presents more training images than in Refs. 15 and 17 and includes the most commonly performed activities in offices, that is, sitting, standing and walking. The software, LabelImg,[30] was used to label all of the images located within both datasets manually. As shown in Figure 2, labels were assigned entirely around each of the specific regions of interest. For some images, more than one occupant appears within the image; hence, multiple labels were assigned.

Table 1. The training and testing images for the occupancy activity detection model.

	Number of images			Number of labels		
Occupancy activity	Training	Testing	Total	Training	Testing	Total
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	

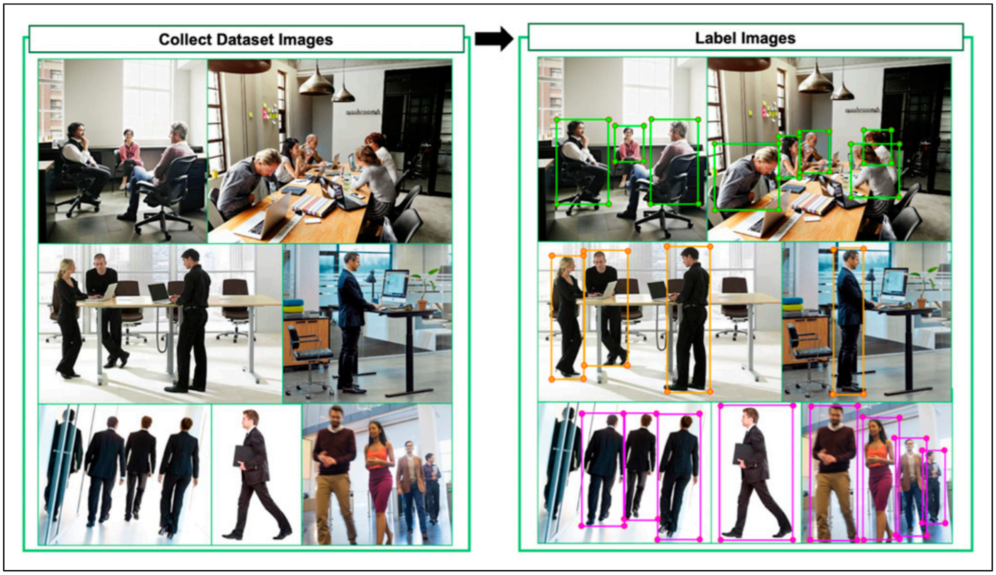


Figure 2. Images of occupants performing common activities within office environments. Examples of how images were labelled. Images obtained from Google.

CNN model selection and configuration. The convolutional neural network is one of the most common types of deep learning algorithms that have been extensively used for image-based classification and recognition tasks with high accuracy.[31] As indicated in the section Datasets and Pre-Processing Stages, input data in the form of images were already collected and labelled. The images can be directly used to derive and extract the acquired features from the selected parts of an image through labelling.[28]

The TensorFlow Object Detection API[32] was selected to help develop the detector. The TensorFlow Object Detection API is a pre-existing open-source framework used to construct, train and deploy deep learning–based models for detection and tracking-based applications. The application of the TensorFlow API is based on a transfer learning approach. This learning method leverages the knowledge learned from one problem by storing the knowledge gained towards the application onto a different but related problem.[33] Effectively, this approach utilises reduced network training time with a smaller input dataset requirement but can still provide an effective detector with high detection performances.

The TensorFlow API framework[34] consists of a range of different pre-trained models that have been successfully trained using common datasets such as the Common Objects in Context (COCO) dataset. This benefitted the training of the detector as the network architecture layers were already defined. For the development of the occupancy detection method, the pre-trained model of the Faster R-CNN with Inception V2 was selected. Faster R-CNN with Inception V2 uses the Faster R-CNN method and Inception V2 architecture directly to classify the type of activity performed within an image or at an instance within a video.[35] Compared to other algorithms such as SSD-MobileNet and YOLO, Faster R-CNN was selected due to its ability to avoid results being dependent on object sizes. The Inception V2 module was used to improve computational performances further while also reducing the computational time required.[32,36] Figure 3 details the architecture and the pipeline configuration of the model used in this present study.

Application of the deep learning model

Initial test based on testing images. Prior to the deployment of the model to a camera for evaluating the

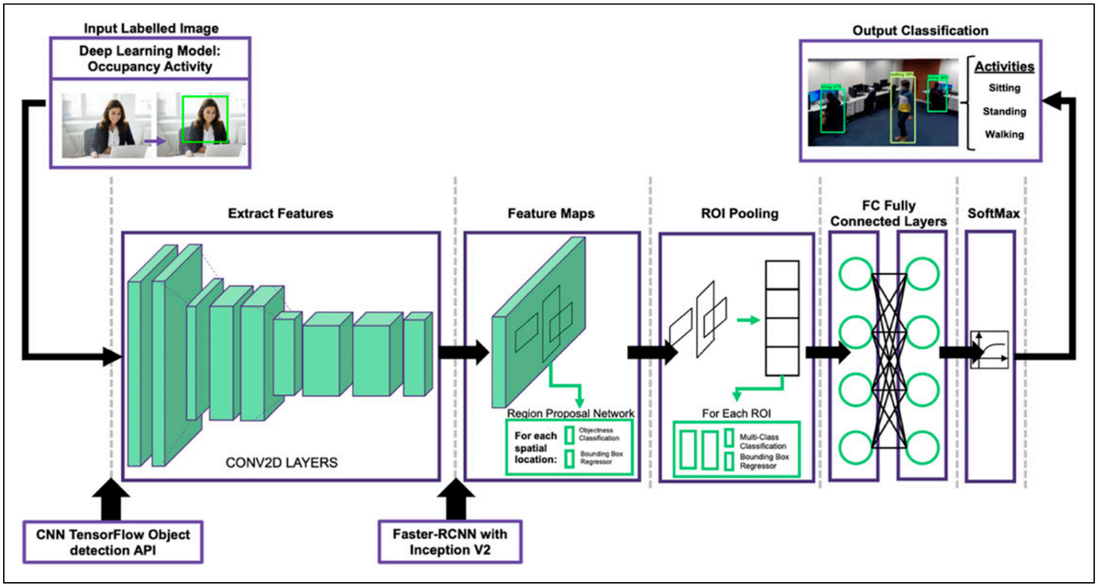


Figure 3. The convolutional neural network (CNN) architecture and configuration used for training the models for occupancy activity detection and recognition.

real-time detection performance, images within the test dataset (Table 1) were used to evaluate the model performance and its ability to classify occupancy activities. Terms of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) were used to assess the performance by providing summaries of the detection and recognition performance of the proposed model in the form of a confusion matrix. Based on these terms, further evaluation was conducted based on the common evaluation metrics given by the following equations for accuracy, precision, recall and F_1 Score. Equations for these metrics are highlighted in equations (1)–(4)

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(P + N)} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$F_1 \text{ Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Case study building. The developed occupancy detection model was tested by conducting an experimental test within an office space on the first floor of the Sustainable Research Building (University of Nottingham, UK). The building consists of a three-storey structure that provides a facility to research sustainable and renewable energy systems. Figure 4(a) presents a photo of the building with the highlighted test room, and Figure 4(b) shows the office space's floor plan. Within the image, the region of the detection area was also identified. It highlighted the possible area captured from the wide field of view. A 1080p resolution camera used during the experimental test. The camera used for detection was positioned at the height and angle replicating typical occupancy sensors, by locating the camera near the room's ceiling. Furthermore, as highlighted in Figure 4(c), three occupants performed a series of activities during the experimental test, and for analysis, the occupants were identified as Person A, B and C.

This case study building was also used to support further analysis of the framework. The building was modelled using the BES tool IESVE[37] to assess the impact of the method on building energy loads. The selected office space has a floor area of 54 m^2 with internal dimensions of 9.24 m (L) × 4.23 m (W), and a

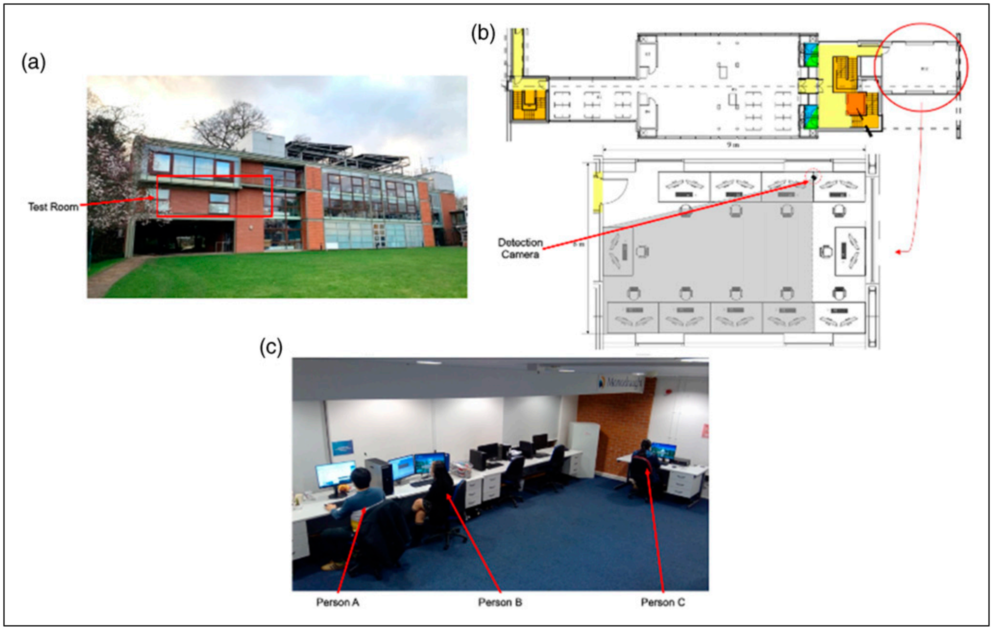


Figure 4. (a) Selected case study building, the Sustainable Research Building at the University of Nottingham, United Kingdom. (b) The layout of the test room selected for the performance of the experimental test. (c) The three occupants indicated as Person A, B and C.

floor-to-ceiling height of 2.5 m (H). The building was divided into different thermal zones, allowing the setup of different operation profiles for each zone. The U-value of the wall was 0.17 W/m^{2}K, the roof was 0.15 W/m^{2}K, the ground was 0.15 W/m^{2}K and glazing was 1.92 W/m^{2}K, which were obtained from the available building drawings. The building is naturally ventilated and integrated with an air-conditioning system. Furthermore, Nottingham, UK weather data file was used for the simulations. Heating profiles were set to maintain indoor temperature at a specific temperature during occupied hours. Details about the BES profiles, including occupancy information, are detailed in the section Test Scenarios and BES.

Real-time detection and DLIP formation. Using the detection model and the selected case study office space, the experimental test was performed for a duration of 15 min during a typical afternoon working day in December. The initial test consisted of three occupants. The occupants during the experimental test were not wearing specified clothing. One male adult and two female adults were wearing typical winter office clothing. The participants carried out the occupancy activities necessary to evaluate the detection method. It should be noted that factors such as age and gender of participants were not considered.

During the experimental test, the detection approach provided real-time predictions of the occupant activities. As observed, each detected and recognised person/activity will have a bounding box that also displays the prediction accuracy above each bounding boxes. Furthermore, data were recorded every second, which was used to form the occupancy count-based DLIP. Figure 5 presents an overview of the formation of the DLIP.

For the formed DLIP to provide valuable information for building control systems or BES, it was coupled with the heat emission rates of occupants performing different activities within an office, as detailed in Table 2.

Test scenarios and BES

Different scenario-based cases were simulated. Each case (Deep Learning Scenario-Based Case 1–4)

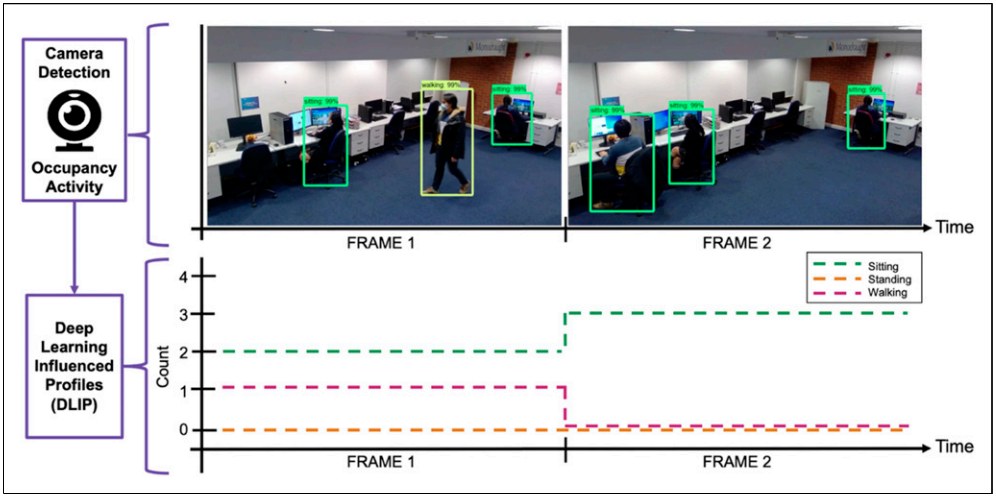


Figure 5. Process of forming the deep learning–influenced profiles from real-time occupancy activity detection and recognition using the deep learning approach.

Table 2. Occupancy activities heat emission rates.[38]

Occupancy activity	Rate of heat emission		
	Total (W)	Sensible (W)	Latent (W)
Sitting	115	75	40
Standing	130	75	55
Walking	145	75	70

represented a different variation in occupancy activity patterns within the office space. These were compared with Typical Office 1 and 2 static profiles, which represents the application of typical, predefined or fixed schedules. The two fixed occupancy profiles are presented in Figure 6(a1) and (a1'). Since there are normally up to eight people working within the office space, a constant number of eight people were assumed during the working hours. Typical Office 1 assumed that the occupants are performing sedentary activities within the office space. While Typical Office 2 assumed a higher activity rate by the occupants. The scenario-based profiles assigned for the Deep Learning Scenario-Based 1, 2, 3 and 4 cases assumed more realistic occupancy patterns within the office space. Description of occupancy patterns for each of these scenarios is detailed under each of the given profiles in Figure 6(a2) to (a5).

Furthermore, heating and cooling profiles were assigned to maintain indoor temperatures within a suitable range to provide occupants' thermal comfort. Current standards and guidelines such as ASHRAE 90.1[39] and ASHRAE 55[40] suggest generalised setpoint temperatures for rooms. During occupied hours, temperature range of 22–27°C was recommended for cooling and 17–22°C for heating. For unoccupied hours, temperatures of 27–30°C for cooling and 14–17°C for heating were also advised. Moreover, CIBSE[41] suggests office buildings to be maintained at an operative room temperature of 21–23°C during the winter and 22–25°C for summer. Hence, all cases were simulated with the room heating setpoint temperature of 21°C–22°C during the building operational hours. In addition, a temperature of 15°C was set for the unoccupied hours, and a cooling setpoint temperature of 25°C were assigned. For both the Typical Office 1 and 2 cases, building operational hours of 06:00–18:00 were assumed. However, for the Deep Learning Scenario-Based cases, the HVAC systems' operation would be based on the detected occupancy level, as indicated in Figure 6(b2) to (b5) and (c2) to (c5). The ventilation profiles (Figure 6(d1) to (d5) were also based on the occupancy patterns.

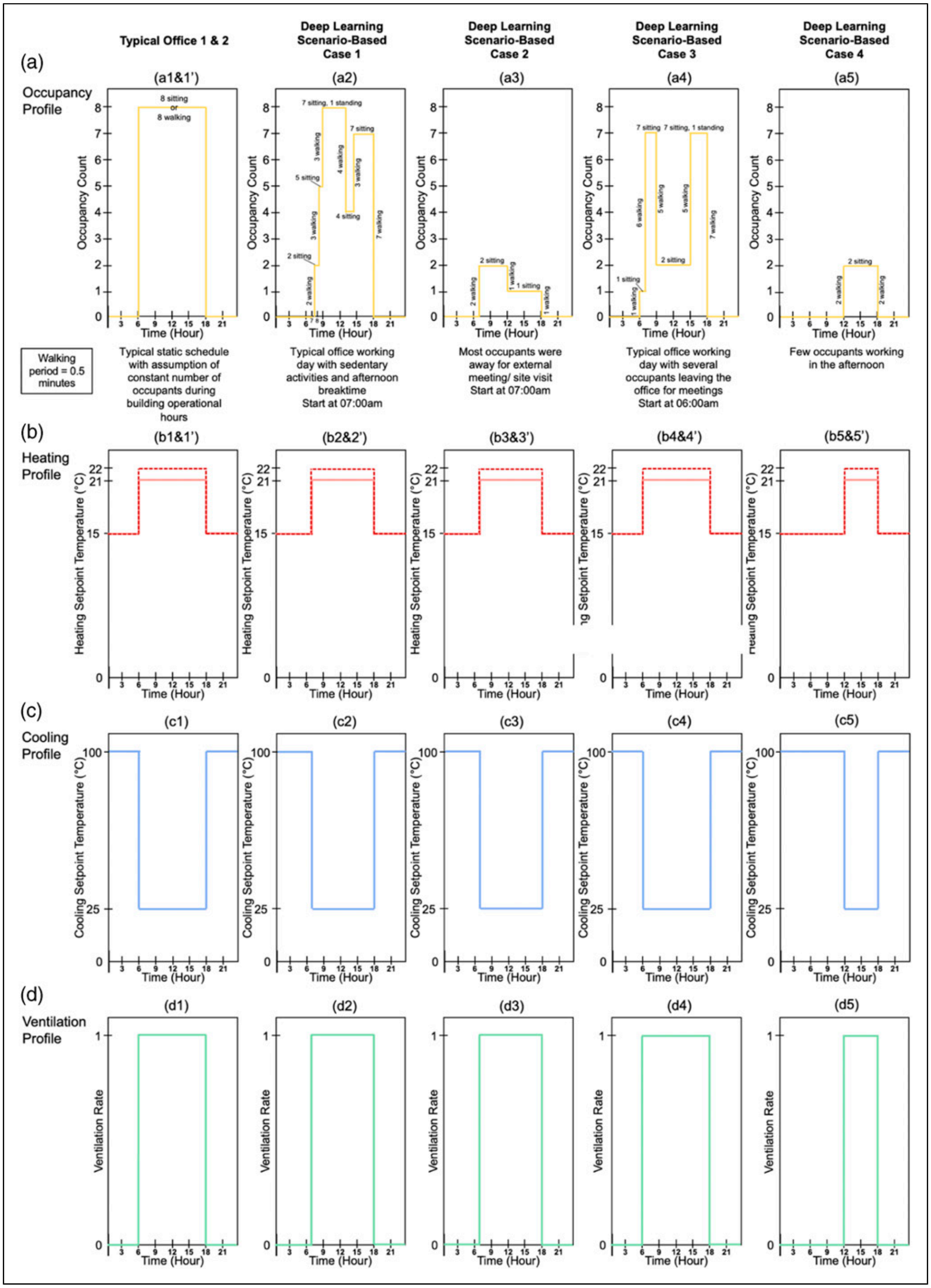


Figure 6. Occupancy and heating, ventilation and air-conditioning profiles for the scenario-based simulation cases (a) occupancy, (b) heating, (c) cooling and d). ventilation profiles for the typical office and the scenario-based simulation cases.

Table 3. Summary of the occupancy and HVAC operation profiles for the simulation of the scenario cases.

	Cases					
	Typical office 1	Typical office 2	Deep learning scenario 1	Deep learning scenario 2	Deep learning scenario 3	Deep learning scenario 4
Description	The deep learning method is not applied Static scheduled profiles assumed		Deep learning occupancy detection model used Heating, cooling and ventilation profiles are dependent on the assigned occupancy profile			
Occupancy profile	Figure 6(a1): Typical Office 1 (constant sitting)	Figure 6(a1'): Typical Office 2 (constant walking)	Figure 6(a2) Deep Learning Scenario-Based profile	Figure 6(a3) Deep Learning Scenario-Based profile	Figure 6(a4) Deep Learning Scenario-Based profile	Figure 6(a5) Deep Learning Scenario-Based profile
Number of occupants present in room	8	8	Varies according to the occupancy profile (achieved by the application of the deep learning detection approach)			
Occupancy internal gains	For sitting: Maximum sensible gain: 70 W/person Maximum latent gain: 45 W/person	Maximum sensible gain: 75 W/person Maximum latent gain: 70 W/person (to meet the maximum total of 145 W/person for the activity of walking)				
Heating profile	Simulation A: Heating setpoint temperature: 21°C during operational hours					
	Figure 6(b1)	Figure 6(b1)	Figure 6(b2)	Figure 6(b3)	Figure 6(b4)	Figure 6(b5)
	Simulation B: Heating setpoint temperature: 22°C during operational hours					
	Figure 6(b1')	Figure 6(b1')	Figure 6(b2')	Figure 6(b3')	Figure 6(b4')	Figure 6(b5')
Cooling profile	Figure 6(c1)	Figure 6(c1)	Figure 6(c2)	Figure 6(c4)	Figure 6(c4)	Figure 6(c5)
Ventilation profile	Figure 6(d1)	Figure 6(d1)	Figure 6(d2)	Figure 6(d4)	Figure 6(d4)	Figure 6(d5)
Ventilation conditions	Infiltration rate: 0.5ACH Maximum conditions: 10 L/s					

Table 3 summarises the building energy simulation modelling profiles and conditions. Each scenario was performed for a day during a typical office week during both heating and cooling seasons.

Results and discussion

Deep learning model training performance and evaluation

The occupancy activity detection model was trained using the graphics processing unit (GPU) NVIDIA GTX1080. Table 4 presents the associated model training results. The training was conducted for 102, 194 steps, and it took 10 h and 29 min for the total losses to reach a converged level. Using the pre-trained model Faster R-CNN with Inception V2 to aid the training of the model, the results provided an average loss of 0.13436 and a minimum loss of 0.005654.

Training performance and model evaluation using test images. Initial detection was performed using the test images within the dataset. This enabled the verification of the model and confirmed that it was successfully trained. The initial recognition performance

Table 4. Deep learning model training results.

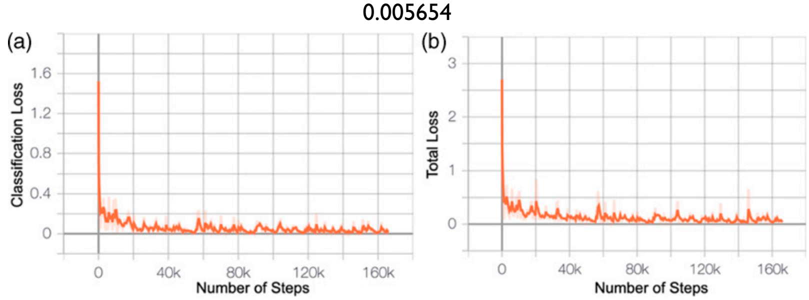
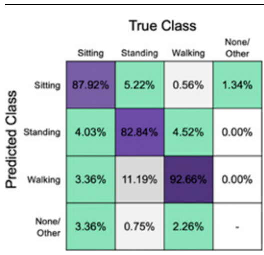
Training conditions and results	Occupancy activity model
Model used	Faster R-CNN with Inception V2
Total steps	102,194
Training duration	10 h 29 min, 52 s
Average loss	0.13436
Minimum loss	0.005654
Training performance	

Table 5. Confusion matrix and model performance results based on the evaluation of the model using the test image dataset.

Confusion matrix	Class	Activity	Accuracy, %	Precision	Recall	F_1 score
	1	Sitting	94.04	0.9250	0.8911	0.9077
	2	Standing	91.43	0.9064	0.8284	0.8657
	3	Walking	92.70	0.8643	0.9266	0.9047

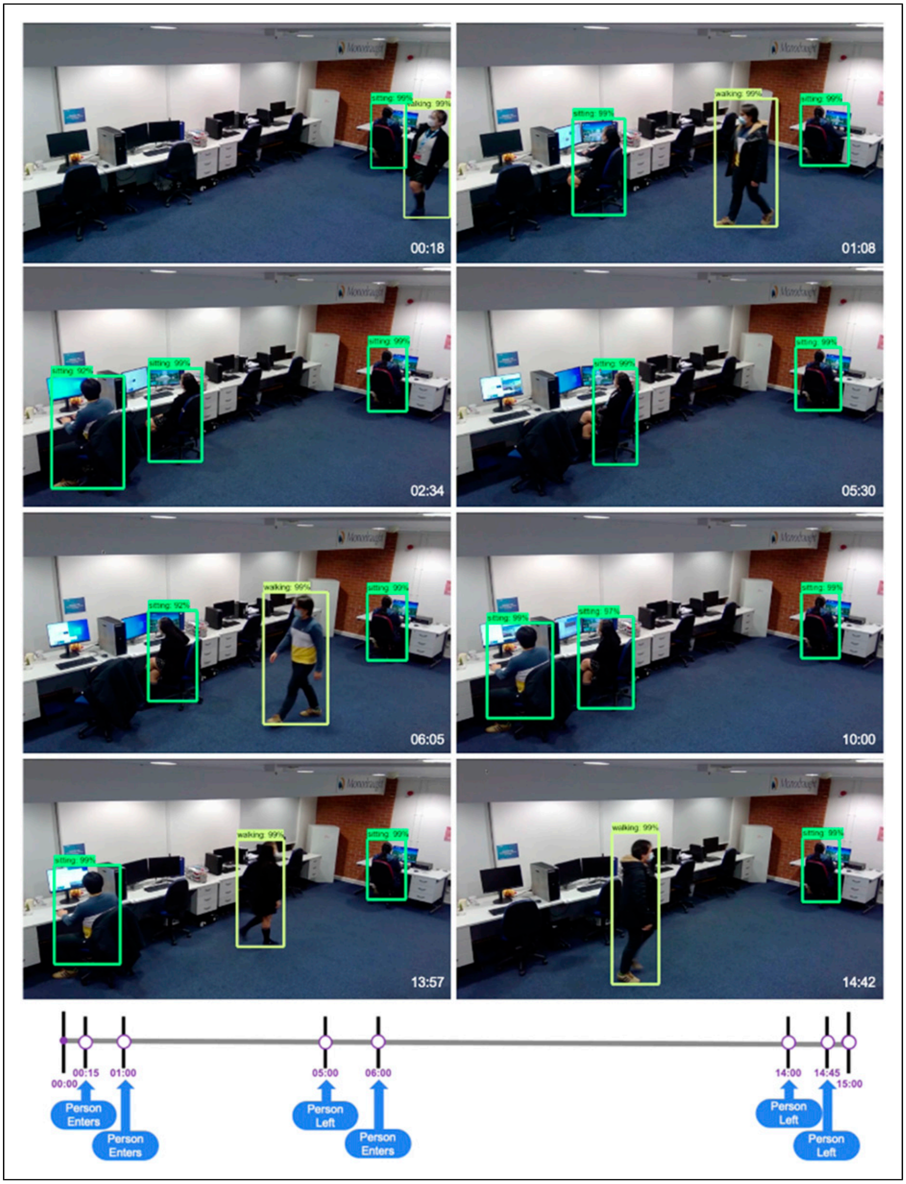


Figure 7. Snapshots of various key stages during the experimental test.

results are detailed in Table 5, with the confusion matrix and the common classification metrics results.

The sitting activity achieved the highest detection accuracy of 94.04%. However, standing and walking activities may have similar occupancy body form and shape, suggesting the possibility of difficulties in identifying the true activity; it therefore resulted in a lower accuracy value of 91.43% for standing and 92.70% for walking. Despite this, the confusion matrix results indicated that most of the prediction labels were

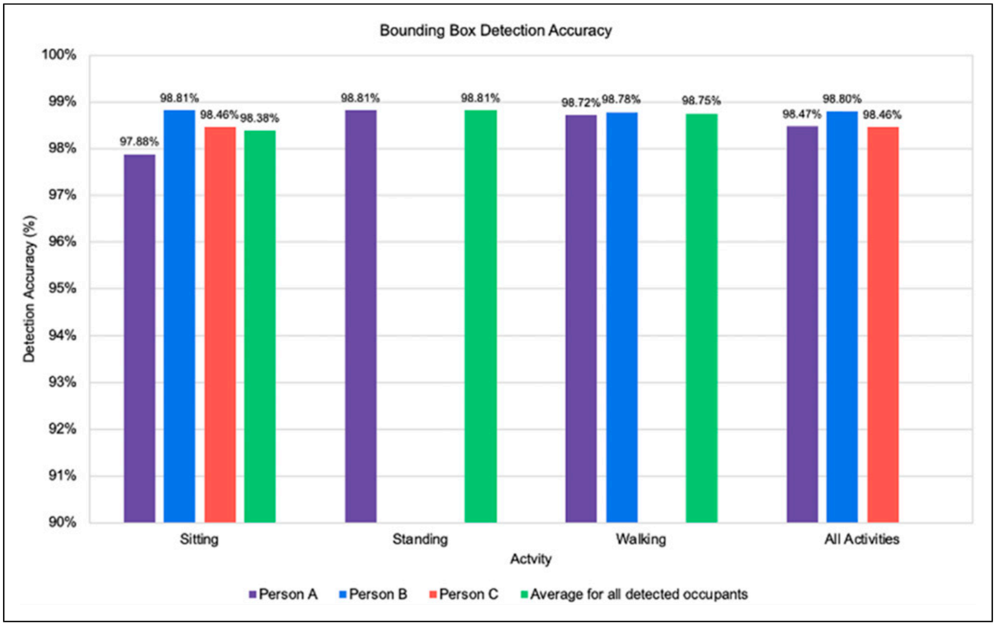


Figure 8. Average detection accuracy based on the Intersection over Union (IoU) values that were generated for the different occupancy activities during the experimental test. These were displayed above each of the generated bounding boxes at every instance when an activity detection was made.

correctly assigned to the desired occupancy activity, showing the potential of the proposed approach.

Framework performance and analysis 1: Detection and recognition performance analysis and profile formation

This section presents the occupancy detection and recognition performance results and analysis based on the experimental test conducted within the case study office.

Real-time detection and recognition of occupancy activities during the experimental test. Based on the experimental test performed within the selected office space, Figure 7 presents a timeline indicating the test's key stages. Example detection images from various stages were presented. It should be noted that in practice, the device will not be storing or outputting images. It will only output real-time information based on the number of occupants performing each activity to generate and establish the deep learning–influenced profiles (DLIPs) of occupancy heat emission.

Analysis of the real-time detection performance. Figure 8 presents the average detection accuracy (this is the detection accuracy values presented above each of the bounding boxes) for each person and activity. It evaluates the detection performance based on the predicted bounding boxes that appeared at every instance in terms of the Intersection over Union (IoU) values generated. Intersection over Union is a standard evaluation metric for CNN detectors used to evaluate how similar a predicted bounding box is to the ground truth box. For such cases, higher prediction accuracy (near to 100%) would be achieved when there is a direct overlap between the target mask and the prediction output. To indicate this, the values presented above each generated bounding box shown in Figure 7 suggests that high confidence is presented for each of the predicted activity, as high percentages were given. As shown in Figure 4(c), for analysis, the

Table 6. The percentage of time the detector achieved correct, incorrect and no/missed detections of the occupants' activities.

	Percentage of detection time achieved		
Activity	Correct detections, %	Incorrect detections, %	No/missed detections, %
Sitting	95.49	0.00	4.51
Standing	38.10	52.38	2.27
Walking	93.18	2.27	4.55
None (no activity)	83.78	16.22	0.00
All activities	93.60	2.18	4.22

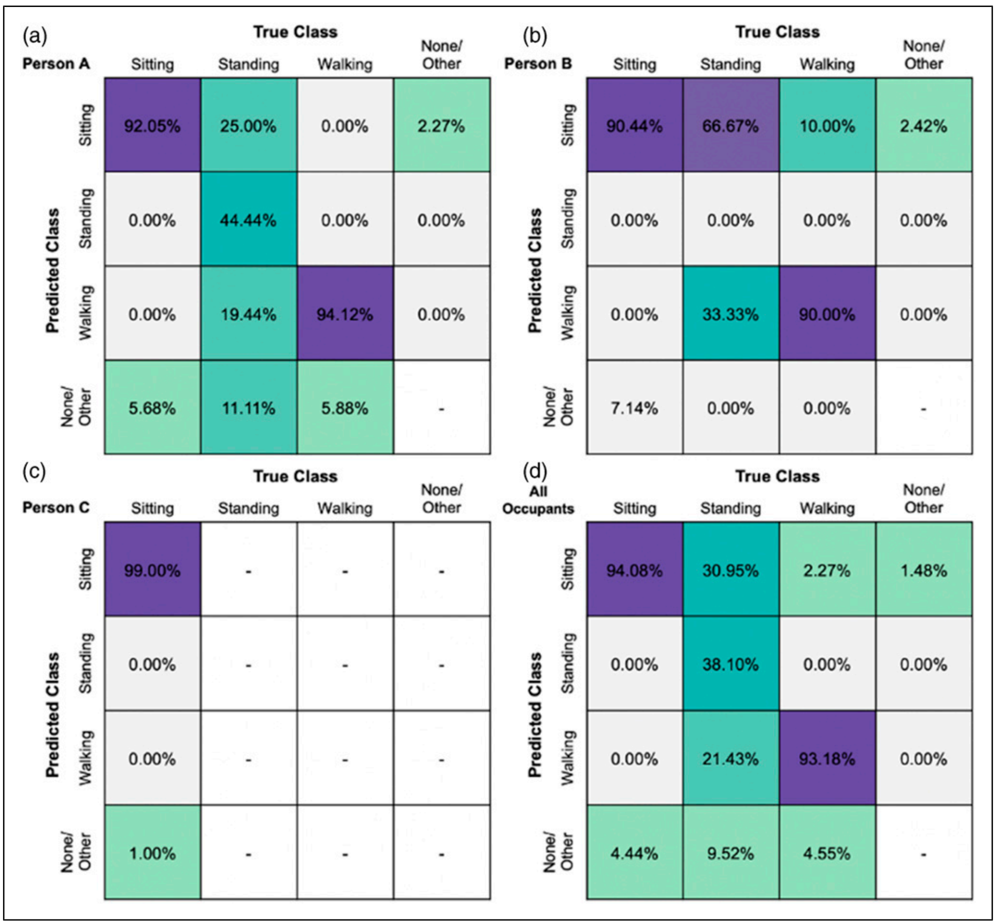


Figure 9. Detection of occupancy activities evaluated in the form of the confusion matrix.

occupants were each given the names Person A, B and C.

Since the experimental test was conducted for 15 min, not all occupants performed all the different activities. However, it still provided enough data to analyse the detection performance. The detection results for Person A, B and C showed up to 98.47%, 98.80% and 98.46% accuracy on average, respectively. The initial results suggest that the distance between the camera and object had a negligible effect on the detection accuracy. However, due to the size of the office room, the influence of further distances cannot be evaluated and should be assessed in future works. Overall, the initial results showed good performance and demonstrated the model's capabilities to recognise the differences between the human poses for each specific activity within an office environment.

Table 6 presents the results for the detection of occupancy activities classified as correct detection, no detection or incorrect detection. The detection data were collected every second. It should be noted that the correct detection was achieved when the activity performed by the person was correctly identified and detection was correctly not made when that activity was not performed. Based on the three activities detected, the best results were achieved for sitting. Due to the similarities of standing and walking poses, the standing activity had incorrect detections for 52.38% of the time, no or missed detections for 9.52% of the time and only correct detection for 38.10% of the time.

Furthermore, to further evaluate the occupancy activity detection performance, the category 'None' was added. This indicated that the model was correct in not identifying that activity was performed when an occupant was absent. Correct detection was achieved 83.78% of the time for this category.

Further evaluation of the detection approach based on classification evaluation metrics. Figure 9 presents the generated confusion matrix for each occupant. This helps display how often the approach classifies one activity as another. The columns represent the predicted activities, and the rows represent the actual or correct activities. This table was then used to calculate the TP, TN, FP and FN, which can then be used to measure the precision, recall and F_1 score as shown in Table 7. Based on the results, the model performed well when detecting and recognising the

Table 7. Evaluation of the occupancy activity detection model performance based on common evaluation metrics.

Class	Activity	Accuracy	Precision	Recall	F_1 score
Person A					
1	Sitting	89.02%	0.7715	0.9419	0.8482
2	Standing	81.48%	1.0000	0.4445	0.6154
3	Walking	81.64%	0.8287	0.9412	0.8814
Person B					
1	Sitting	71.26%	0.5335	0.9268	0.6772
2	Standing	66.67%	—	0.0000	0.0000
3	Walking	85.56%	0.7297	0.9000	0.8060
Person C					
1	Sitting	99.00%	1.0000	0.9900	0.9950
2	Standing	—	—	—	—
3	Walking	—	—	—	—
All occupants					
1	Sitting	86.95%	0.7305	0.9549	0.8278
2	Standing	79.37%	1.0000	0.3810	0.5518
3	Walking	90.58%	0.8130	0.9318	0.8684
Average for all activities		85.63%	0.8478	0.7559	0.7493

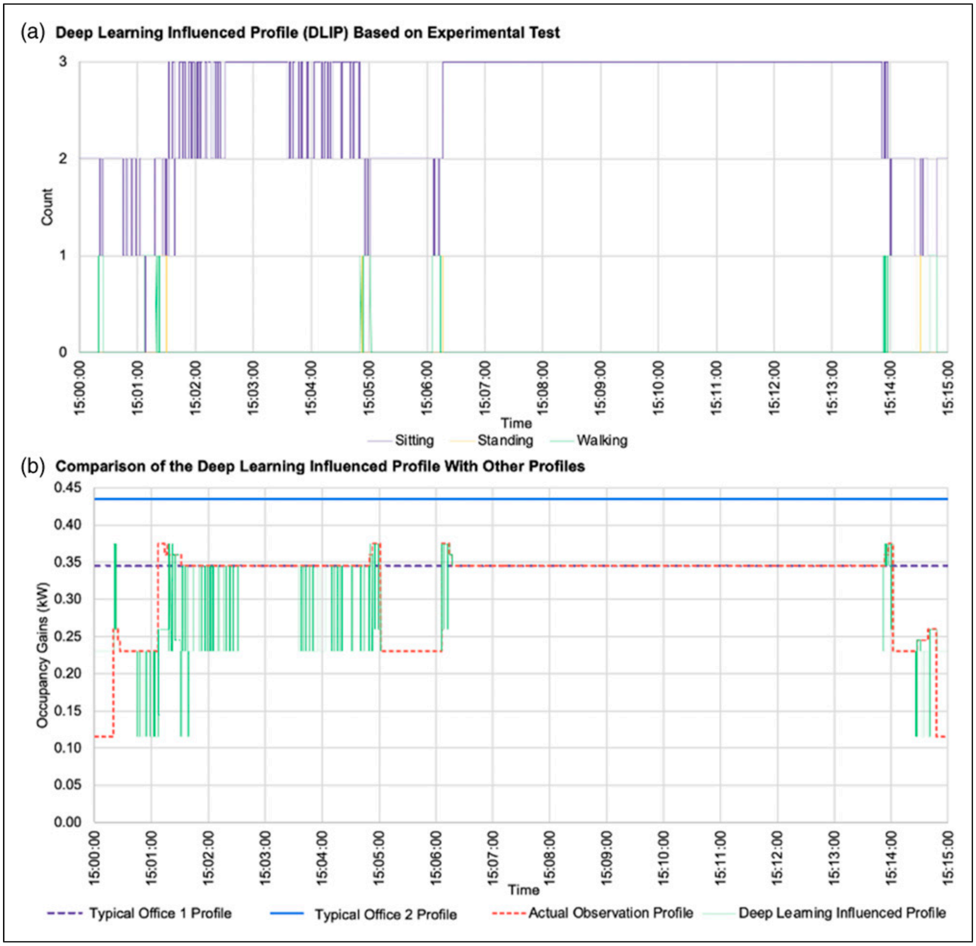


Figure 10. (a) Occupancy activity profiles generated by the proposed detection method. (b) Generated occupancy heat emission profiles plotted against Typical Profiles and Actual Observation Profile.

sitting activity, with the occasional prediction of the category none or other. In addition, there were times when sitting activity was predicted even when an occupant was not present. These were considered incorrect detections as it recognised some of the chairs within the detection space as occupants sitting.

However, the detection and recognition of the standing activity were the poorest. As identified by Figure 9, it suggests that there was an approximately equal split in achieving correct detection, with it being identified as sitting, and also it being identified as either walking or not being identified at all. However, the detection of the walking activity showed good performance. The distribution of the results attained for the different responses based on the different factors, including the angle, distance and position in relation to the detection camera, shows the approach's ability to achieve good detection performance. However, the initial results also showed that the current model has limitations and should be further developed.

Occupancy activity-based deep learning–influenced profiles. The approach does not require the data to collected and stated in the form of images or videos to avoid privacy issues. Instead, the process presented in Figure 5 was used to generate deep

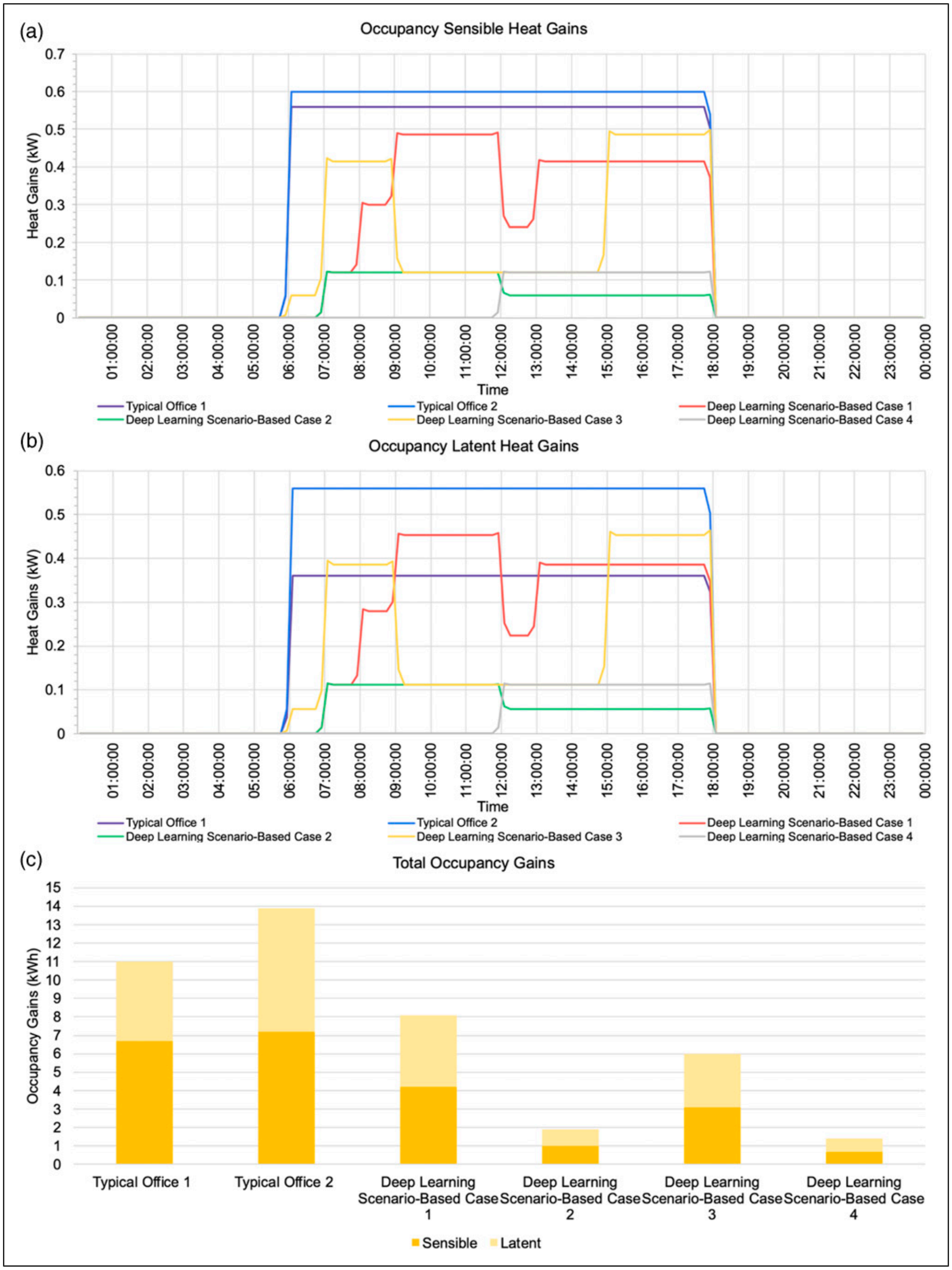


Figure 11. Comparison of the occupancy heat gain profiles generated using the proposed approach and the typical occupancy scheduled profiles. Variation of gains across time: (a) occupancy sensible heat gains, (b) occupancy latent heat gains and (c) the total occupancy gains.

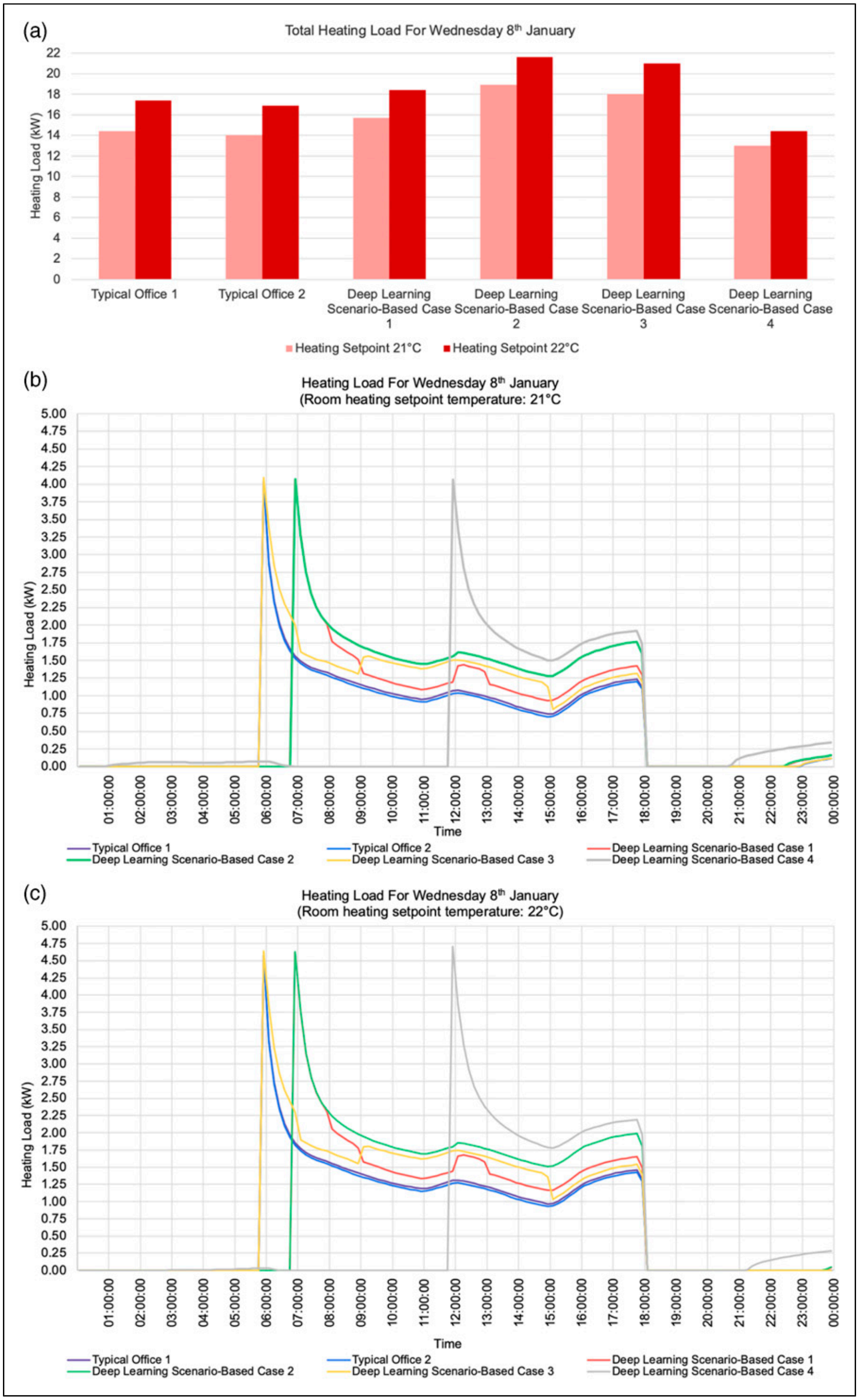


Figure 12. (a) Comparison of the total heating load achieved under the different deep learning scenario-based cases compared to the typical occupancy profiles, with room heating setpoint at 21°C and 22°C. Variation of heating load across time for all simulated cases when a room heating setpoint temperature of (b) 21°C and (c) 22°C was assigned.

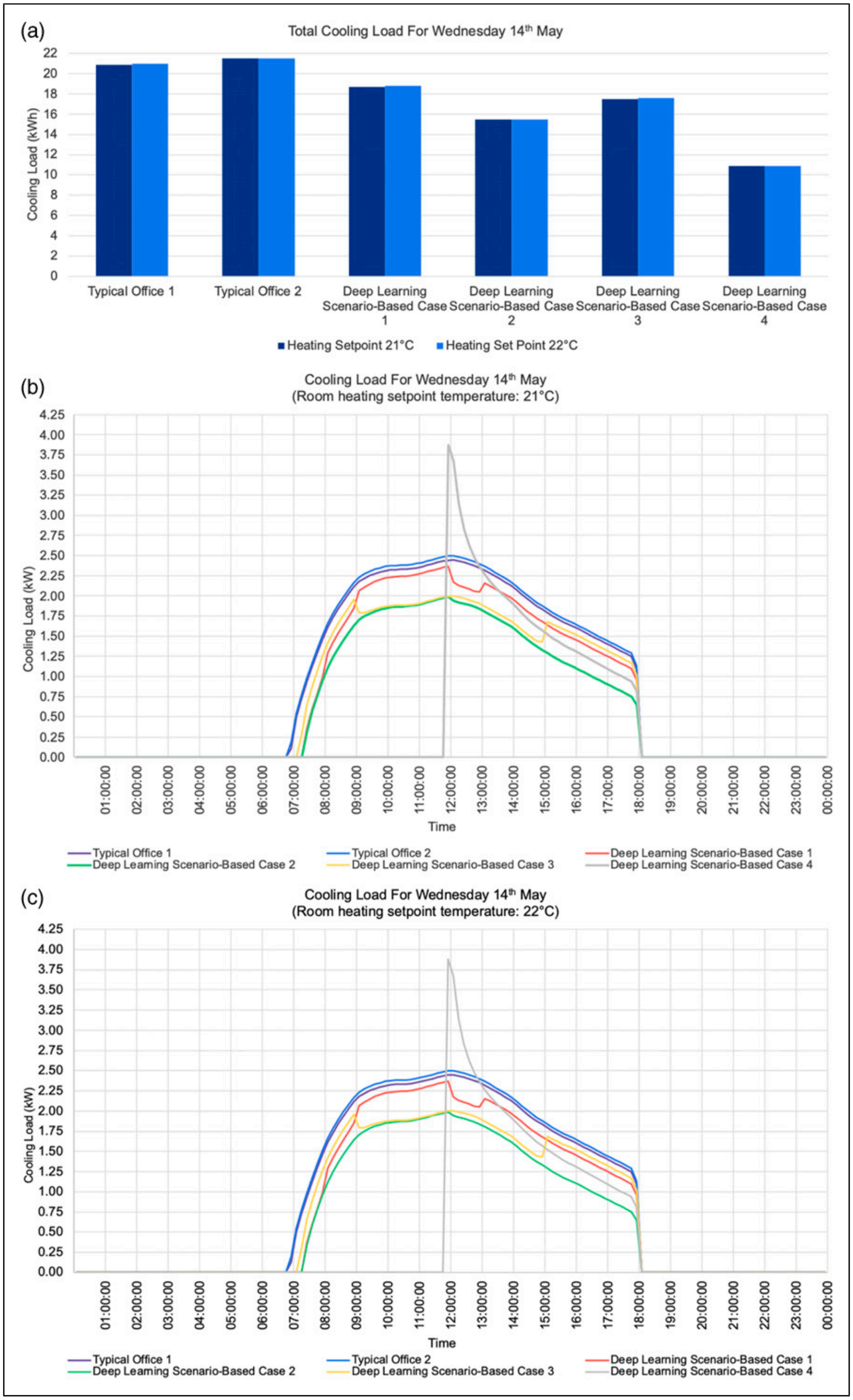


Figure 13. (a) Comparison of the total cooling load achieved under the different deep learning scenario-based cases compared to the typical occupancy profiles, with room heating setpoint at 21°C and 22°C. Variation of cooling load across time for all simulated cases when a room heating setpoint temperature of (b) 21°C and (c) 22°C was assigned.

learning–influenced profiles (DLIPs) of the occupancy activities as shown in Figure 10(a). Furthermore, the DLIPs were then converted to form an occupancy heat emission–based profile which can be used to inform and assist the operations of building system controls and for BES models.

Figure 10(b) shows a comparison of the DLIP with other profiles such as the Typical Office 1 and 2 profiles. The profiles' results indicate that the u of the DLIP predicted 29.09% and 10.60% lower occupancy heat gains as compared to Typical Profiles 1 and 2. However, the results also showed that the DLIP still has errors when compared with the Actual Observation Profile. This represents the 'true' activities performed by the occupants. Overall, it had an error of 4.14%.

Framework performance and analysis 2: Scenarios and building energy performance analysis

Occupancy heat gains. Based on the cases highlighted in Table 3 and the occupancy profiles given in Figures 6(a) and 11 presents the predicted daily occupancy sensible and latent heat gains. Typical Office 1 and 2 results showed the predicted occupancy heat gains when static or fixed occupancy profiles were set. Clearly, the use of such profiles can lead to unrealistic and less diverse variation in occupancy heat gains. The four Deep Learning Scenario-Based cases indicate that occupancy gains were directly related to the number of occupants and the type of activities performed by each occupant during the day. For Case 4, the room was unoccupied most of the time, and only a small number of occupants were present for a few hours. This led to the lowest occupancy heat gains (1.40 kWh). However, if such a situation happens and the building was operated based on the assumption corresponding to the occupancy profiles for Typical Office 1 or 2, it can lead to an overestimation by up to 9.60 kWh and 12.50 kWh.

Heating load. The following results present the heating demand for the simulation cases for a typical winter day, Wednesday 8th January. Given that the simulations were performed with two different heating setpoint temperatures during the building operational hours, Figure 12(a) shows that heating load for all cases would increase by an average of 2.62 kWh when the heating setpoint was set at 22°C as compared to 21°C. Therefore, changing the room set point temperature by 1°C can change the total heating demand by 15.40%. By depending on the occupancy level, slight adjustments to the setpoint can be done to significantly affect the energy demand of the building.

Figure 12(b) and (c) present the distribution of the predicted daily heating load. The results in Figure 12(a) showed that the Deep Learning Scenario-Based cases had a higher heating load than Typical Office 1 and 2. It indicated that the use of the detection approach led to a reduction in occupancy heat gains, which impacted the heating load. Case 2 achieved the highest heating load with 18.90 kWh and 21.60 kWh for the two setpoint temperatures. This was due to the low number of occupants present and lower emission rate activities (mostly sitting), resulting in low occupancy heat gains. In comparison, Case 1 had the highest amount of occupancy heat gains and heating loads of 15.7 kWh and 18.4 kWh. While the lowest heating demand was observed for Case 4, which had no occupants in the morning and hence the heating setpoint was automatically lowered to 15°C when using the proposed control strategy.

Cooling load. Figure 13 presents the cooling loads for the various simulation cases during a typical day in the cooling season, Wednesday, 14th May. Since the cooling setpoint temperature was maintained at 25°C during building operational hours (Figure 6), Figure 13(a) suggests the changes in the heating setpoint of 21°C and 22°C presented minimal variations towards the total cooling loads, with only an increase by an average of 0.05 kWh, which is equivalent to a change of 0.26%.

As shown, the total cooling loads were 20.90 kWh, and 21.00 kWh for Typical Office 1, and 21.50 kWh for Typical Office 2. Overall, using typical profiles for occupancy would result in higher cooling loads than the deep learning approach–based cases. Higher amounts of cooling are required when there is a higher occupancy heat gain in the space. For example, Deep Learning Scenario 1 achieved the highest occupancy gains. This resulted in the highest cooling load with 18.80 kWh.

Like the evaluation for heating, the lowest cooling loads were achieved by Case 4 as the deep learning detection approach assisted the cooling system's operations. Based on the response to the detections made, the cooling setpoint temperature becomes 25°C once occupants were present within the space. Hence, no cooling was required prior to 12:00, and only after this time, cooling up to a peak of 3.879 kW was required to provide a thermally comfortable environment for occupants.

Conclusion and future work

This study presents the analysis of the application of a vision-based deep learning approach for occupancy activity detection conducted within an open-plan office space. The detection model was first developed by using a transfer learning-based method to establish and train a CNN for the classification of occupancy activities. The model was then deployed towards a camera which enabled the performance of real-time detections. The initial performance of the model was evaluated based on a 15-min experimental detection test. Average detection accuracy of 98.65% was achieved across all activities. During the real-time detection of a selected space, constant data about the number of occupants performing each of the selected activities were generated as the DLIP. The generated DLIP was compared to typical static profiles indicating the ability to provide more valuable data about occupants within a building space, regarding heat emission for more effective HVAC system operations.

Building energy simulation was performed with various scenario-based cases to assess the deep learning approach and to provide insights into how the proposed detection method can enable HVAC systems to adapt and respond to occupancy's dynamic changes. Results indicate that the deep learning approach can reduce the under- or over-estimation of occupancy heat gains. Compared with static profiles, when occupancy activities are monitored, it influences occupancy gains and ultimately results in a greater variation in the building energy demands. Results in terms of both heating and cooling loads were highly dependent on the occupancy profiles. Using the deep learning approach through the scenario-based cases suggests the heating, cooling and ventilation profiles assigned would operate based on the response to the detections made to reduce unnecessary building energy loads effectively.

Improvements to the current approach are required prior to the integration with controls of building HVAC systems. Future works include developing a streamlined framework-based solution to define the required HVAC control system conditions based on real-time detection data responses. This includes the most suitable indoor/room setpoint temperature that should be assigned to HVAC systems to provide adequate thermal conditions based on the real-time understanding of the utilisation of the space by occupants.

Acknowledgements

The authors would also like to thank the participants for their support in conducting the field experiments. The present publication was first presented at the CIBSE ASHRAE Technical Symposium 2020, 14–15 September 2020 with the paper entitled 'Energy management and optimisation of HVAC systems using a deep learning approach'. The authors would also like to thank CIBSE ASHRAE for awarding this paper as the 'Most significant contribution to the art & science of building services engineering'.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the University of Nottingham, and the PhD studentship from EPSRC, Project Reference: 2100822 (EP/R513283/1).

ORCID iDs

Paige Wenbin Tien https://orcid.org/0000-0003-0123-248X
Shuangyu Wei https://orcid.org/0000-0002-1053-1652
John Kaiser Calautit https://orcid.org/0000-0001-7046-3308

References

1. Intergovernmental Panel on Climate Change. Contribution of working group III to the fifth assessment report of the intergovernmental panel on climate change. Report of the Intergovernmental Panel on Climate Change. In: AR5 *Climate change 2014: Mitigation of climate change*. Cambridge, New York: Cambridge University Press, 2014.
2. IEA. *Transition to sustainable buildings, strategies and opportunities to 2050*. Paris, France: International Energy Agency, 2013.
3. Yang L, Yan H and Lam JC. Thermal comfort and building energy consumption implications - A review. *Appl Energy* 2014; 115: 164–173.
4. Clarke M-A and CIBSE Chartered Institution of Building Services Engineer. One vision: integrating intelligent HVAC. *CIBSE J* 2020: 30–31.
5. Tien PW, Calautit JK, Darkwa J, et al. A deep learning framework for energy management and optimisation of HVAC systems. In: IOP conference series: earth and environmental science, Bangkok, Thailand, 11–14 December 2019.
6. Wei S and Calautit JK. Development of deep learning-based equipment heat load detection for energy demand estimation and investigation of the impact of illumination. *Int J Energy Res* 2020; 45(5): 7204–7221.
7. Hong T, Yan D, D'Oca S, et al. Ten questions concerning occupant behavior in buildings: the big picture. *Build Environ* 2017; 114: 518–530.
8. Tahmasebi F and Kang J and CIBSE Chartered Institution of Building Services Engineer. Research on indoor air quality for homeworkers. *CIBSE J* 2020: 24–25.
9. Peng Y, Rysanek A, Nagy Z, et al. Occupancy learning-based demand-driven cooling control for office spaces. *Build Environ* 2017; 122: 145–160.
10. Annaqeeb MK, Markovic R, Novakovic V, et al. Non-intrusive data monitoring and analysis of occupant energy-use behaviors in shared office spaces. *IEEE Access* 2020; 8: 141246–141257.
11. Sun B, Luh PB, Jia QS, et al. Building energy management: integrated control of active and passive heating, cooling, lighting, shading, and ventilation systems. *IEEE Trans Autom Sci Eng* 2013; 10: 588–602.
12. Shih H-C. A robust occupancy detection and tracking algorithm for the automatic monitoring and commissioning of a building. *Energy Build* 2014; 77: 270–280.
13. Burak Gunay H, O'Brien W and Beausoleil-Morrison I. Development of an occupancy learning algorithm for terminal heating and cooling units. *Build Environ* 2015; 93(Pt 2): 71–85.
14. Labeodan T, Zeiler W, Boxem G, et al. Occupancy measurement in commercial office buildings for demand-driven control applications-A survey and detection system evaluation. *Energy Build* 2015; 93: 303–314.
15. Tien PW, Wei S, Calautit JK, et al. A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand-driven control solutions. *Energy Build* 2020; 226: 110386.
16. Wei S, Tien PW, Calautit JK, et al. Vision-based detection and prediction of equipment heat gains in commercial office buildings using a deep learning method. *Appl Energy* 2020; 277: 115506.
17. Tien PW, Wei S, Calautit J, et al. Occupancy heat gain detection and prediction using deep learning approach for reducing building energy demand. *J Sustain Dev Energy Water Environ Syst* 2020: 1080378.
18. Ramanan D and Forsyth DA. Finding and tracking people from the bottom up. In: Proceedings/CVPR, IEEE computer society conference on computer vision and pattern recognition, Madison, WI, 18–20 June 2003. .
19. Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image eecognition. arXiv 2014 arXiv:1409.1556.
20. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the 2014 IEEE conference on computer vision and pattern recognition, Columbus, OH, 24–27 June 2014.
21. Yuan L, Qu Z, Zhao Y, et al. A convolutional neural network based on TensorFlow for face recognition. In: Proceedings of the 2017 IEEE 2nd advanced information technology, electronic and automation control conference (IAEAC), Chongqing, China, 25–26 March 2017.
22. Zou J, Zhao Q, Yang W, et al. Occupancy detection in the office by analyzing surveillance videos and its application to building energy conservation. *Energy Build* 2017; 152: 385–398.
23. Markovic R, Grintal E, Wölki D, et al. Window opening model using deep learning methods. *Build Environ* 2018; 145: 319–329.

24. Pérez-Hernández F, Tabik S, Lamas A, et al. Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: application in video surveillance. *Knowl Based Syst* 2020; 194: 105590.
25. Seema S, Goutham S, Vasudev S, et al. Deep learning models for analysis of traffic and crowd management from surveillance videos. In: H Das, P Pattnaik, S Rautaray, et al. (eds) *Progress in computing, analytics and networking*. Singapore: Springer, 2020, pp. 83–93.
26. Xu C and Chen H. A hybrid data mining approach for anomaly detection and evaluation in residential buildings energy data. *Energy Build* 2020; 215: 109864.
27. Dhillon A and Verma GK. Convolutional neural network: a review of models, methodologies and applications to object detection. *Prog Artif Intelligence* 2019; 9(2): 85–112.
28. Rawat W and Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput* 2017; 29(9): 2352–2449.
29. Sultana F, Sufian A and Dutta P. A review of object detection models based on convolutional neural network. In: J Mandal and S Banerjee (eds) *Intelligent computing: image processing based applications*. Singapore: Springer; 2020, pp. 1–16.
30. Tzutalin LabelImg. Git code. https://github.com/tzutalin/labelImg (2015, accessed 13 February 2021).
31. O'Mahony N, Campbell S, Carvalho A, et al. Deep learning vs. traditional computer vision. In: K Arai and S Kapoor (eds) *Advances in computer vision. CVC 2019. Advances in intelligent systems and computing*. Cham: Springer; 2019, pp. 128–144.
32. Huang J, Sun C, Zhu M, et al. Speed/accuracy trade-offs for modern convolutional object detectors, 2016. arXiv:1611.10012.
33. Shin H-C, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016; 35(5): 1285–1298.
34. TensorFlow. Tensorflow detection model zoo, https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf1_detection_zoo.md (2020, accessed 13 February 2021).
35. Wu H, Liu Q and Liu X. A review on deep learning approaches to image classification and object segmentation. *Comput Mater Continua* 2019; 60(2): 575–597.
36. Pachidis TP, Sarafis IT and Lygouras IN. Real time feature extraction and standard cutting models fitting in grape leaves. *Comput Electron Agric* 2020; 74(2): 293–304.
37. IES. IESVE integrated environmental solutions, https://www.iesve.com/ (2019, accessed 13 February 2021).
38. CIBSE Chartered Institution of Building Services Engineer. *Table 6.3 typical rates at which heat is given off by human being in different states of activity. Environmental design: CIBSE guide A*. London: CIBSE, 2015.
39. ASHRAE. Energy standard for buildings except low-rise residential buildings. ANSI/ASHRAE/IES Standard; 90: 1–2019.
40. ASHRAE. Thermal environmental conditions for human occupancy. Standard 55.
41. CIBSE Chartered Institution of Building Services Engineer. *Table 1.5 recommended comfort criteria for specific applications. Environmental design: CIBSE Guide A*. London: CIBSE, 2015.