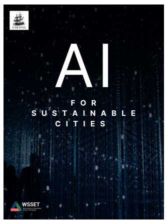
Artificial Intelligence for Sustainable Cities | Volume 1 (2026) | ISSN 2979-1464

CARAVEL

Publish open access at Caravel Press

Artificial Intelligence for Sustainable Cities

Journal homepage: aiforsustainablecities.com



Technical article

AI-Based Occupancy Detection for HVAC-Relevant Load Estimation in a University Building

Rufus Dwamena [1], Wuxia Zhang [1], Paige Tien [1], John Calautit [1] and Hao Sun [1,2]*

[1] *Department of Architecture and Built Environment, University of Nottingham, United Kingdom.*
[2] *Global Centre for Clean Air Research (GCARE), University of Surrey, United Kingdom.*

ABSTRACT

Occupancy-aware HVAC control depends on realistic estimates of occupant-related heat gains, yet many buildings still operate on static schedules that assume sustained high occupancy. This study evaluated whether a vision-based people detector could generate a more representative occupancy schedule for a university seminar room. Two YOLOv7 models were trained on 500 annotated images, including five null images, and deployed to two simultaneous 60-minute recordings of room B5 in the Marmont Centre, University of Nottingham, from front-right and back-left viewpoints. Held-out validation favoured the extended Google Colab configuration over the Roboflow configuration, with mAP, precision and recall of 88.4%, 89.3% and 85.8% versus 86.9%, 84.1% and 80.9%, respectively. Against minute-by-minute manual counts, the best configuration - Google Colab with the front-right view - achieved a normalized count accuracy of 95.8%. Occupancy counts were translated into occupant heat-gain schedules using CIBSE benchmark data. The best dynamic schedule yielded a reconstructed mean occupant gain of 2.37 kW compared with 4.48 kW for a static seminar-room benchmark. Within the original EnergyPlus comparison framework, the best dynamic schedule corresponded to normalized winter-heating and summer-cooling indices of 73.3 and 32.4 relative to a static benchmark of 100, equivalent to reductions of 26.7% and 67.6%. The results show that camera placement and occlusion are as important as model choice and that vision-based occupancy detection can materially improve HVAC-relevant schedule realism in teaching spaces, provided that both counting accuracy and the direction of counting error are validated.

ARTICLE INFO

Article History:
Received: 19 April 2026
Revised: 08 June 2026
Accepted: 29 June 2026

Keywords:
Occupancy detection
HVAC control
Computer vision
YOLOv7
EnergyPlus
Seminar room

Article Citation:
Dwamena, R., Zhang, W., Tien, P., Calautit, J., & Sun, H. (2026). AI-Based Occupancy Detection for HVAC-Relevant Load Estimation in a University Building. Artificial Intelligence for Sustainable Cities, 1(1), 163–174. https://doi.org/10.65582/aifsc.2026.011

1. INTRODUCTION

Buildings remain central to decarbonization because the energy used to heat, cool, ventilate and operate occupied space accounts for a substantial share of total building-related energy demand and carbon emissions. Recent global assessments show that the buildings and construction sector still accounts for roughly one third of global energy use and energy-related carbon emissions, which means that relatively modest operational inefficiencies become important at system scale (UNEP and GlobalABC 2025). A persistent weakness in conventional control is the reliance on fixed schedules or conservative design assumptions about occupancy.

* Corresponding author. Email address: hao.sun@surrey.ac.uk (Hao Sun)

In many educational and commercial buildings, HVAC systems are operated for long periods as if rooms were fully, or near-fully, occupied throughout the day, even though real attendance fluctuates sharply across hours, days and activities. Occupant-centric control seeks to reduce this mismatch by responding to the people actually present rather than to a conservative timetable (Naylor et al. 2018; Pang et al. 2023; Wang et al. 2022; Wang et al. 2023b).

For HVAC operation, occupancy matters for more than simple presence detection. The number of people in a room directly influences sensible heat gains, latent moisture gains, contaminant generation and the amount of outdoor air required to maintain acceptable indoor air quality. Guidance used in building design therefore expresses internal heat gains and ventilation demand partly on a per-person basis (CIBSE 2021; UK Government 2021). In a seminar room, underestimating headcount risks inadequate ventilation and local discomfort, while overestimating it can lead to unnecessary heating, fan energy or mechanical cooling. The quality of the occupancy input is therefore fundamental. A binary occupied/unoccupied signal may be sufficient for switching lights, but robust HVAC control benefits from knowing how many people are present and how that count changes over time (Tien et al. 2020; Tien et al. 2021c; Wei et al. 2022b).

Traditional sensing approaches such as passive infrared and CO_2 monitoring remain useful because they are relatively low cost and easy to integrate, but they also have well-known limitations when the aim is real-time headcount rather than simple presence detection. PIR sensors depend on motion and can miss seated or still occupants; CO_2 sensors provide an indirect signal that is affected by mixing, lag, infiltration and prior room history. Reviews of occupancy detection technologies consistently note that no single conventional sensor type offers a complete solution for fine-grained room-level counting across varied building types (Chen et al. 2018; Rueda et al. 2020). Vision-based methods are therefore attractive because a camera can, in principle, observe both the presence and the number of occupants directly, and earlier work has already shown that video-derived counts can be combined with environmental measurements to inform building control (Wang et al. 2017; Tien et al. 2020; Tien et al. 2021a; Tien et al. 2021c; Wei et al. 2022b).

Among vision-based approaches, deep learning object detectors are especially relevant because they can provide rapid inference from ordinary RGB images and can be deployed on practical camera feeds (Wei et al. 2020; Tien et al. 2021a; Tien et al. 2021b; Pincott et al. 2022; Tien et al. 2022b). The detector architecture used in this study was YOLOv7, a real-time one-stage detector that remains a strong baseline for fast visual inference when speed and accuracy must be balanced (Wang et al. 2023a). Yet the choice of detector is only part of the problem. In real classrooms and seminar rooms, counting performance depends strongly on viewpoint, occlusion, furniture layout, backlighting and whether occupants are seen frontally, obliquely or from behind (Zhang et al. 2024). Much of the literature evaluates detectors using generic detection metrics alone, whereas HVAC applications require a different question: does a given model-view combination produce a room-count schedule that is accurate enough to alter the inferred internal gains and ventilation demand in a meaningful way (Tien et al. 2020; Tien et al. 2021c; Wang et al. 2022; Wei et al. 2022a; Zhang et al. 2024)? These visual factors are not incidental. Real teaching spaces present variable lighting, including daylight ingress, glare and dim lecture-mode conditions, as well as hard occlusion in which occupants may be partly or wholly hidden behind structural elements such as pillars, projector screens or the people in front of them. Both effects can suppress detections and therefore distort the recovered count, so they are treated here as first-order determinants of schedule quality rather than as edge cases.

This study addresses that applied gap in a teaching-space case study. Two YOLOv7 detectors, developed through separate training pipelines, were deployed to two simultaneous views of the same seminar room so that both algorithmic performance and camera placement could be assessed under identical occupancy dynamics. The detected counts were converted into minute-by-minute occupancy schedules, compared against manual ground truth, translated into HVAC-relevant sensible and latent gains, and then interpreted through an EnergyPlus-based comparison against a conventional static benchmark (Tien et al. 2020; Tien et al. 2021c; Wang et al. 2022; U.S. Department of Energy 2026). The central research question was straightforward: which model-view combination yields the most reliable room-count schedule, and how different is the resulting HVAC-relevant load profile from the static schedule typically assumed in design and operation? The novelty of the case study lies in linking room-level counting performance directly to load reconstruction rather than treating object detection accuracy as an end in itself. Although the empirical work is deliberately scoped to a single room, the detector-to-schedule-to-load workflow is intended as a transferable building block for scalable smart-campus management, in which room-level occupancy estimates feed building- and campus-wide supervisory control, digital-twin models and autonomous building-management systems (Bortolini et al. 2022; Tien et al. 2022; Wang et al. 2023b).).

2. MATERIALS AND METHODS

2.1. CASE-STUDY ROOM AND RECORDINGS

Room B5 in the Marmont Centre at the University of Nottingham was selected as the case-study space because it represents a realistic teaching-space control problem rather than an idealized laboratory setting. The room has a floor area of 95.28 m^2 and a nominal capacity of 38 seats, and it is regularly used for seminar-based teaching (Figure 1). This makes it a suitable environment for testing occupancy-aware HVAC logic because attendance is neither constant nor predictable: the room passes through an arrival phase, a settled teaching phase, and a departure phase, all of which produce different internal-load conditions over a relatively short period. In practical terms, it is exactly the kind of space in which fixed occupancy assumptions can lead to over-conditioning for much of the operating period.

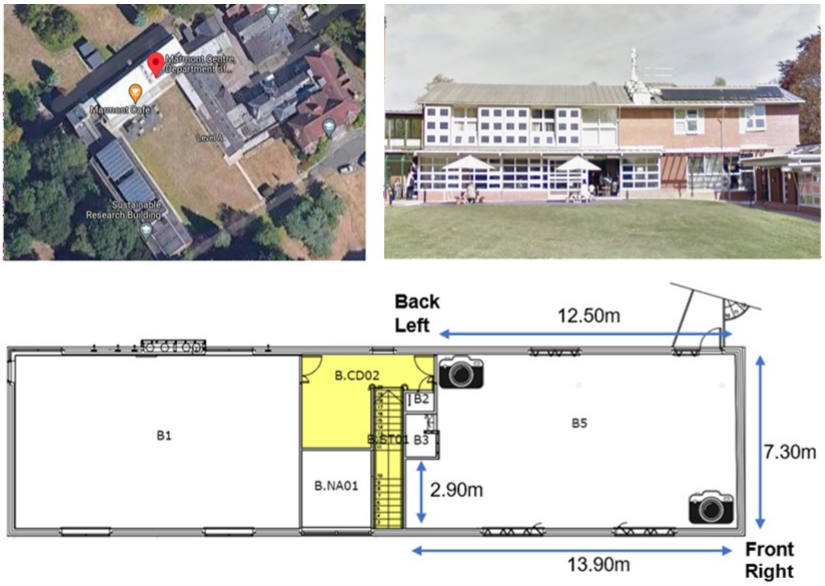


Figure 1. Site and room context for the Marmont Centre case study: location of the Marmont Centre on the University of Nottingham campus; plan context of room B5 showing the case-study space and the approximate back-left and front-right camera positions used for simultaneous recording.

Two identical 1080p cameras mounted on tripods recorded the same 60-minute seminar simultaneously from opposite corners of the room (Figure 2). One camera was positioned to obtain a back-left (BL) view and the other to obtain a front-right (FR) view. Using two synchronized recordings of the same event allowed the effect of viewpoint to be tested without changing the occupants, teaching activity, or timing of arrivals and departures. Any difference in the recovered occupancy schedule could therefore be interpreted mainly as a consequence of camera placement, occlusion, and field of view rather than differences in room use.

Figure 2. **(a)** Back-left camera view and **(b)** front-right camera view used for occupancy detection. The two cameras recorded the same seminar from opposite corners of room B5, allowing the influence of viewpoint, field of view, and occlusion on count accuracy to be assessed directly.

Manual ground-truth counts were produced at one-minute intervals from each recording. Ground truth was defined with respect to the visible field of the specific camera rather than by assuming that either camera had perfect visual coverage of the room. This is a deliberate methodological choice. The two views have different blind spots, different lines of sight to seated occupants, and different degrees of overlap between foreground and background people. Evaluating each detector against its own visible scene is therefore fairer and more operationally meaningful than comparing both cameras with an idealized whole-room count that neither camera could reliably observe. In an HVAC application, this distinction matters because deployed systems must work with the view they are given, not with perfect geometric coverage.

The room context is also relevant for interpreting the results. The seminar layout creates several counting challenges that are typical of live educational settings: seated occupants are partly hidden by desks and by the shoulders and heads of people in front of them; movement near the start and end of the session temporarily changes visibility; and some people sit close to the image edge, where partial truncation is more likely. For this reason, the case study does not simply test whether a detector can recognize people in a generic classroom image; it tests whether a detector can recover a room-count schedule that remains useful for HVAC interpretation under realistic visual constraints.

Table 1. Case-study, training and HVAC-translation assumptions used in the study.

Parameter	**Value**
Case-study room	Room B5, Marmont Centre, University of Nottingham; 95.28 m^2; 38 seats
Recordings	Two simultaneous 1080p recordings of one 60-minute seminar from back-left and front-right corners
Training data	500 annotated images, including 5 null images, with a 70/20/10 train/validation/test split
Static benchmark	47 W/m^2 for seminar rooms (27 W/m^2 sensible + 20 W/m^2 latent)
Dynamic gain conversion	130 W/person (93 W sensible + 37 W latent) (CIBSE Guide A 2021; metabolic rate consistent with ASHRAE 55-2020)
Environmental basis	19C winter, 25C summer; 10 L/s/person

2.2. MODEL DEVELOPMENT AND HELD-OUT VALIDATION

A single class, occupant, was annotated across the image set because the downstream task was room-level counting rather than identity recognition or fine-grained activity classification. The dataset contained 500 annotated images, including null images with no occupants present, and deliberately spanned a range of crowd sizes, viewpoints, and room conditions. This variety was intended to reduce over-specialization and to expose the detector to both sparse and dense scenes. The dataset was divided using a 70/20/10 split for training, validation, and held-out testing so that model development and final image-based evaluation were separated.

Two YOLOv7 training pipelines were then assessed. The first used Roboflow, a cloud-based annotation-and-training service whose free tier restricted training to 300 epochs. The second used Google Colab, a cloud-based GPU training environment in which the same annotated dataset was trained for up to 1000 epochs so that training could continue beyond the Roboflow limit and the loss curve could be allowed to stabilize more fully. Using the same annotations and the same train/validation/test split in both cases meant that the comparison focused on the practical consequences of the training pipeline and training duration rather than on differences in the underlying data. YOLOv7 was chosen because it remains a strong real-time object-detection baseline for fast visual inference in applications where near-live processing is desirable (Wang et al. 2023). It should be emphasised that the Roboflow-versus-Colab contrast is an infrastructural comparison rather than a comparison of detection algorithms. In effect a constrained cloud annotation-and-training service capped at 300 epochs versus an extended cloud-based GPU training environment run to 1000 epochs. Holding the architecture, annotations and data split constant isolates the effect of training budget and pipeline, but it does not benchmark YOLOv7 against alternative architectures. Comparison with two-stage detectors such as Mask R-CNN, together with the practical trade-off between cloud-based GPU training and local edge-inference at deployment, is therefore identified as a priority for future work rather than claimed here.

The held-out metrics in Table 2 show that both pipelines produced capable detectors, but the Google Colab model performed better across all three reported indicators: mAP, precision, and recall. That pattern is important even before room deployment because counting performance depends heavily on both missed detections and false detections. A model with higher recall is less likely to miss partially visible occupants, while a model with higher precision is less likely to misclassify background objects or ambiguous image regions as people. Even modest differences at this stage can accumulate into meaningful errors in the recovered occupancy schedule when predictions are repeated over an entire seminar.

Table 2. Case-study, training and HVAC-translation assumptions used in the study.

Model	Training regime	mAP (%)	Precision (%)	Recall (%)
Roboflow	300 epochs (platform limit)	86.9	84.1	80.9
Google Colab	1000 epochs	88.4	89.3	85.8

At the same time, the held-out results should be interpreted as an initial benchmark rather than as a sufficient test of HVAC usefulness. Still-image validation establishes that the model has learned the person-detection task, but a detector that performs well on held-out images may still behave poorly in a real seminar if camera position, seated posture, overlap between occupants, or lighting conditions differ from the training examples. For that reason, the image-based metrics in Table 2 are used here as a precursor to the more important room-count evaluation in Section 2.3 and the real-room results in Section 3.1.

2.3. ROOM-COUNT EVALUATION AND HVAC TRANSLATION

Real-room performance was assessed by comparing predicted room counts with manually produced counts at each one-minute time step. Four complementary metrics are reported: mean absolute error (MAE), root mean square error (RMSE), bias, and normalized count accuracy. MAE indicates the typical absolute size of the counting error in persons, while RMSE places greater weight on larger misses and is therefore useful for identifying models that fail disproportionately during high-occupancy periods. Bias indicates whether a detector tends to under-count or over-count systematically, which is especially important for building control because persistent under-counting and persistent over-counting have different operational consequences.

Normalized count accuracy is defined as

$$100 \times \left[1 - sum\left(\left|N_{hat_t} - N_t\right|\right) / sum(N_t)\right] \tag{1}$$

where N_t is the manual count and N_{hat_t} is the detected count at time step t. This metric provides a schedule-level measure of fidelity by relating total absolute count error to the total occupancy observed over the full period. In the present study, it is particularly useful because HVAC control depends on the quality of the whole time-varying schedule rather than on isolated frames. The metric therefore complements MAE and RMSE by indicating how closely the detector reproduces the overall occupancy profile that would be supplied to a supervisory control strategy.

To interpret the results in HVAC terms, each dynamic schedule was converted into an occupant heat-gain schedule using CIBSE Guide A data for seated, moderate office work at 20 °C: 93 W sensible and 37 W latent per person, giving 130 W/person in total (CIBSE 2021). This conversion reflects the fact that headcount influences more than simple presence status. The number of people in a room affects sensible heat release, latent moisture addition, and the ventilation requirement that must be maintained to dilute contaminants and maintain acceptable indoor air quality (Chen et al. 2018; Rueda et al. 2020). A dynamic count schedule is therefore more relevant to HVAC than a binary occupied/unoccupied signal when the goal is to estimate internal gains and fresh-air demand. The per-person figures used here are fixed benchmark coefficients and so do not capture occupant-to-occupant variation in metabolic rate (activity level) or clothing insulation. For seated, light seminar activity the CIBSE Guide A value is broadly consistent with the metabolic ranges tabulated in ASHRAE Standard 55 (ASHRAE 2020). A fuller thermal-load and comfort analysis that incorporates metabolic and clothing variability, for example through the PMV/PPD framework of ASHRAE 55, is identified as a refinement for future work.

A static benchmark schedule was also defined so that the dynamic schedules could be compared against a conventional design-and-operation assumption. Using seminar-room allowances from CIBSE Guide A, the benchmark was set at 27 W/m^2 sensible plus 20 W/m^2 latent, equivalent to 4.48 kW for room B5. The environmental assumptions adopted for the simulation were a 19 °C winter setpoint, a 25 °C summer setpoint. Because all test cases used the same room, the same broad operating assumptions, and the same simulation framework, differences between scenarios can be attributed primarily to differences in the occupancy schedule rather than to differences in geometry or plant setup.

EnergyPlus was retained as the building-simulation environment because it provides a transparent and reproducible framework for comparing the load implications of different occupancy schedules. However, to avoid overstating numerical precision where only the relative effect of schedule change could be recovered robustly, the heating and cooling results are presented as normalized indices with the static schedule fixed at 100 rather than as unverifiable raw plant values. This preserves the central comparison while keeping the analysis methodologically cautious. In effect, the simulation is used here to answer a comparative question: how different is the HVAC-relevant load profile when static occupancy assumptions are replaced by a detector-derived dynamic schedule? (U.S. Department of Energy 2026). The end-to-end data pipeline therefore proceeds through six explicit stages (Figure 3a): (i) raw video capture from each camera; (ii) one-minute frame sampling; (iii) YOLOv7 occupant detection and bounding-box counting to obtain the per-minute count N_t; (iv) conversion to a fractional occupancy schedule N_t/N_{design}; (v) multiplication by the per-person sensible and latent gains to populate the EnergyPlus People object; and (vi) comparative load simulation reported as normalised indices. In the present study this mapping was performed by a scripted post-processing step rather than by a live software bridge. Replacing that manual step with an automated middleware layer or API that streams counts directly into the building-simulation or building-management system would make the integration fully reproducible and is a clear direction for deployment-oriented work. It should also be acknowledged that the simplified single-zone geometry limits the absolute fidelity of the simulation, which is one reason the results are reported as relative indices and why validation against in-situ indoor-air-quality sensing, for example CO_2 and temperature logging, is recommended before operational use.

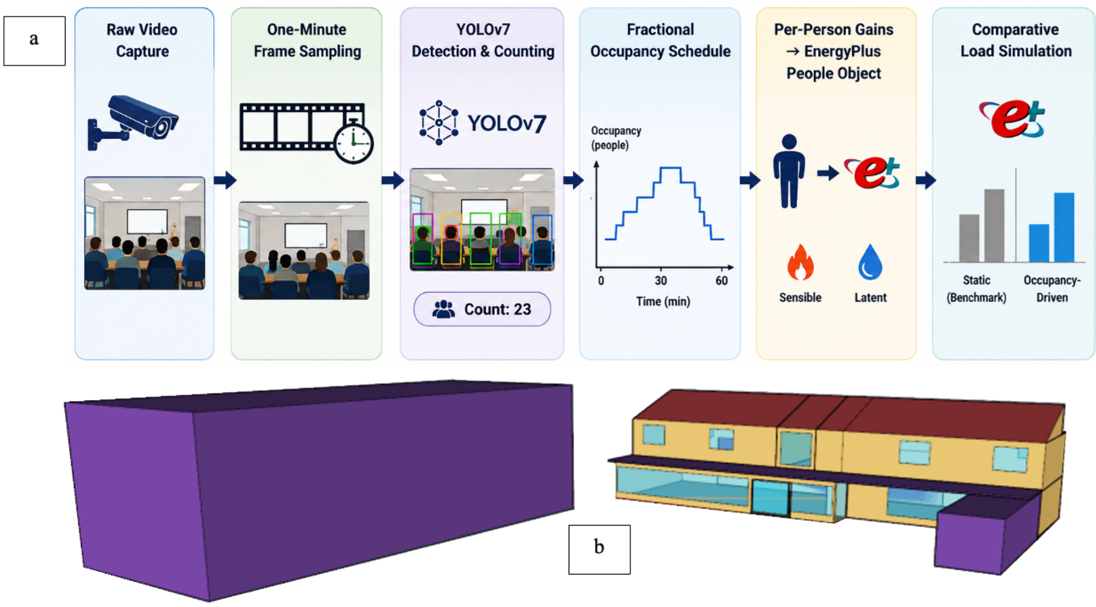


Figure 3. **(a)** Data pipeline linking vision-based occupancy detection to HVAC-relevant load simulation. **(b)** Simplified geometry model used for the EnergyPlus comparison. Room B5 was treated as the focal thermal zone for occupancy-schedule testing, with the time-varying occupant gain computed as Q(t) = N(t) x 130 W/person (93 W sensible + 37 W latent), while adjacent building context and neighbouring massing were retained where necessary to represent boundary conditions and external shading.

3. RESULTS

3.1. VIEW-DEPENDENT OCCUPANCY COUNTING

The held-out metrics in Table 2 already favour the Google Colab model, but the more important question is how the detectors behave in a live seminar room. Figure 4 shows the minute-by-minute occupancy profiles recovered from the two views. All four detector-view combinations captured the overall temporal structure of the seminar reasonably well: occupancy rose during arrival, remained high through the main teaching period, and then fell rapidly at the end of the session. This broad agreement is reassuring because it shows that even the weaker detector recovered the main phases of room use rather than producing an arbitrary or unstable signal.

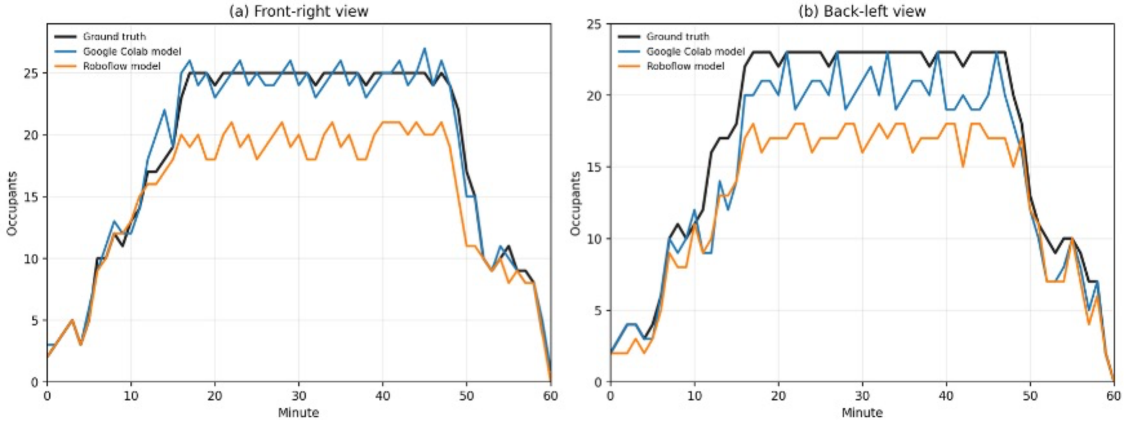


Figure 4. **(a)** Front-right and **(b)** back-left occupancy profiles for the two detectors against manual ground truth. The Google Colab model follows the manual count more closely, while the Roboflow model shows consistent under-counting during the occupied plateau.

Table 3. Real-room counting performance against manual minute-by-minute occupancy counts.

View	Model	MAE (persons)	RMSE (persons)	Bias (persons)	Normalized count accuracy (%)
Front-right	Roboflow	3.26	4.11	-3.20	82.0
Front-right	Google Colab	0.75	1.09	+0.07	95.8
Back-left	Roboflow	3.87	4.47	-3.87	76.5
Back-left	Google Colab	1.90	2.46	-1.87	88.5

The decisive difference emerges during the high-occupancy plateau, where count accuracy matters most for HVAC interpretation. This is the period in which ventilation demand is highest and in which internal gains are most likely to diverge from static assumptions. Here the Google Colab detector remained much closer to the manual counts, especially in the front-right view. By contrast, the Roboflow model substantially under-counted during the plateau in both views. The pattern suggests that the Roboflow detector struggled more when the room was full, occupants were seated, and bodies were partially hidden by desks and by other people in the foreground. In other words, the lower-performing model was weakest exactly when the occupancy signal mattered most.

The difference between the two camera positions is also instructive. The front-right view produced the most HVAC-useful schedule because it captured more of the seated audience and suffered less severe occlusion of visually informative upper-body features. Its manual count accumulated 1106 occupant-minutes with a peak visible occupancy of 25, compared with 1005 occupant-minutes and a peak of 23 from the back-left view. The front-right Google Colab configuration was nearly unbiased, with a mean predicted count of 18.20 persons versus a manual mean of 18.13. The back-left Google Colab configuration, although still substantially better than the Roboflow model, retained a negative bias of -1.87 persons. This is an important distinction: a detector that systematically misses people may produce a lower-looking load schedule, but that lower schedule is not necessarily a more truthful one.

These results reinforce the point that viewpoint is not a minor implementation detail. In room-level occupancy detection, the camera is part of the sensing model. A detector that performs well in one view can produce a materially less reliable schedule in another view of the same space because the visibility of seated occupants changes with angle, depth stacking, and edge-of-frame truncation. For HVAC control, this matters because a negatively biased schedule can translate into under-conditioning and under-ventilation. That risk is especially significant in teaching spaces, where both thermal comfort and fresh-air provision depend on how many people are actually present.

Figure 5. Sample real-world detections from the seminar-room recordings: front-right example frame and back-left example frame. The bounding boxes illustrate successful occupant recognition under live teaching conditions while also highlighting the main sources of residual error, including seated posture, partial occlusion, overlap between occupants, and edge-of-frame truncation. These residual-error sources are quantified by the view-dependent accuracy in Table 3: in the less-occluded front-right view the best model achieved an MAE of 0.75 persons, about 4% of the mean occupancy, with negligible bias (+0.07), whereas in the more heavily occluded back-left view the same model showed an MAE of 1.90 persons and a systematic under-count of -1.87 persons. The error margin attributable to occlusion in this configuration therefore corresponds to roughly two missed occupants on average during the occupied plateau.

3.2. HVAC-RELEVANT LOAD IMPLICATIONS

The reconstructed mean occupant gains are shown in Figure 6a. The static benchmark produces 4.48 kW because it assumes a fixed seminar-room allowance across the full period regardless of how many people are actually present at any given time. By comparison, the front-right ground-truth schedule yields 2.36 kW and the front-right Google Colab schedule yields 2.37 kW, an almost exact match. This agreement is important because it shows that the best detector-view combination was not merely good in abstract counting terms; it reproduced the HVAC-relevant gain profile closely enough that the manually derived and detector-derived schedules are effectively interchangeable at the level of mean occupant load.

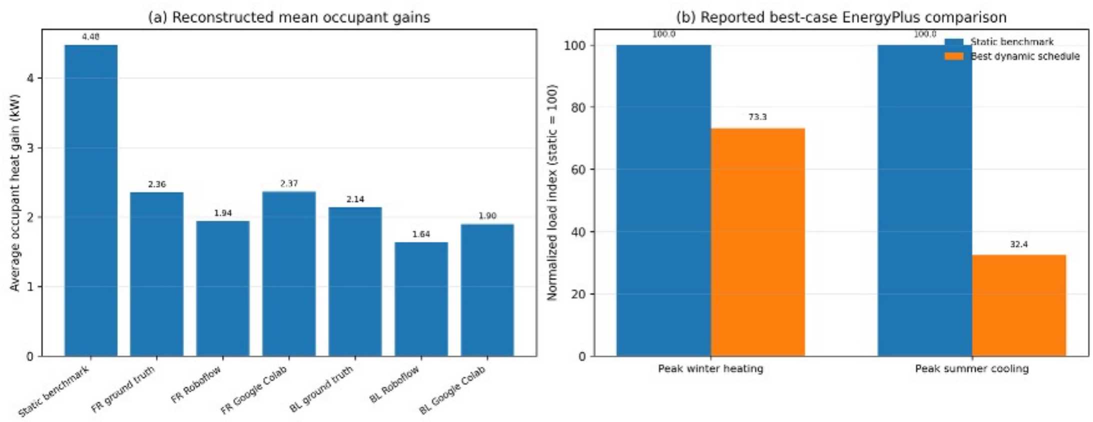


Figure 6. **(a)** Reconstructed mean occupant gains derived from the occupancy schedules and **(b)** normalized load indices with the static benchmark fixed at 100. The lowest apparent loads are not automatically the most accurate loads; view-related under-counting can make a schedule look efficient while simply missing people.

Relative to the static benchmark, that best dynamic schedule is 47.2% lower. This reduction is physically plausible and operationally meaningful. The seminar was not occupied continuously at its design maximum, and attendance rose and fell within the hour. A static schedule based on a full seminar-room allowance therefore overstated occupant-related gains for much of the period. The value of the dynamic approach is not that it produces a low number as such, but that it produces an estimate better aligned with the room's actual use. In this case, the dynamic estimate is lower because real attendance was lower and more variable than the benchmark assumption. Because the occupant-gain reconstruction scales linearly with the counted headcount at 130 W/person, the sensitivity of the load estimate to detection error can be bounded directly: a uniform plus or minus 10% count error propagates to an approximately plus or minus 10% change in reconstructed occupant gain, i.e. roughly plus or minus 0.24 kW about the 2.37 kW best-case estimate. Off-peak periods, where headcount and therefore gains are small, are correspondingly less sensitive in absolute terms, whereas the peak plateau, where ventilation demand is highest, is where a given percentage error has the largest absolute effect on inferred load and comfort. This bounding argument is analytical; a full coupled sensitivity study that propagates the measured error distribution through EnergyPlus across peak and off-peak hours remains future work.

The back-left Google Colab schedule yields an even lower reconstructed gain of 1.90 kW, but this should not be interpreted as superior performance. The same view also showed a stronger negative counting bias, which means part of the apparent gain reduction is simply missed occupancy. This is a central interpretive point of the paper: an apparently efficient schedule is not automatically a more accurate schedule. In control applications, the distinction matters because underestimating people in the room may suppress heating, cooling, or ventilation demand on paper while producing poorer real conditions for occupants.

Figure 6b summarizes the corresponding normalized EnergyPlus comparison. With the static case fixed at 100, the best dynamic schedule corresponds to a winter-heating index of 73.3 and a summer-cooling index of 32.4. Relative to the static benchmark, these values imply reductions of 26.7% for peak winter heating and 67.6% for peak summer cooling. The larger reduction in cooling than in heating is consistent with the role of internal gains in room conditioning. In summer, excess assumed occupancy directly adds to the cooling burden, whereas in winter internal gains can partially offset heating demand. Overestimating attendance therefore tends to penalize cooling more strongly than heating.

Taken together, the results show that the benefit of dynamic occupancy schedules depends on both realism and trustworthiness. A validated dynamic schedule can reduce the gap between assumed and actual room loads and thereby support more appropriate HVAC operation. A biased dynamic schedule, by contrast, can appear beneficial while actually reflecting sensing error. The correct interpretation is therefore not simply that "dynamic is better than static", but that validated dynamic occupancy is better than static, and that validation must include both counting accuracy and the direction of error.

4. DISCUSSION

The main contribution of the study is that it evaluates vision-based occupancy detection in terms that are directly meaningful for building operation. Many computer-vision studies stop at image-level detection metrics, whereas HVAC applications require something more specific: a reliable time-varying estimate of how many people are in the room and how that estimate changes the internal-load picture seen by the building system. In that respect, the present case study shows that model choice and camera placement are both first-order determinants of schedule quality. The detector cannot be assessed independently of the view from which it operates, because the useful signal is produced jointly by the learned model and the visual geometry of the room.

The front-right Google Colab configuration performed best because it combined the more extensively trained detector with the more informative room view. It was nearly unbiased relative to manual counting and reproduced the manually derived occupant gain almost exactly. The back-left Google Colab configuration was still usable, but its stronger negative bias illustrates the risk of treating all camera positions as equivalent. In a live control context, that bias would not merely lower a statistical score; it could lead directly to underestimation of the ventilation requirement and of occupant-related heat and moisture gains. The study therefore supports a practical design principle: camera siting should be treated as part of HVAC-sensing design rather than as an afterthought once the detector has been trained.

The comparison with the static benchmark further explains why occupancy-aware control remains attractive in decarbonization-oriented building operation. In this seminar, peak attendance was brief rather than continuous, and the room spent much of the hour below its visible maximum occupancy. A fixed benchmark schedule therefore overstated occupant-related loads for most of the period. That result aligns with the broader occupant-centric control literature, which shows that building systems can reduce energy use when control decisions are tied more closely to real occupancy conditions rather than to fixed timetables or conservative design assumptions (Naylor et al. 2018; Pang et al. 2023; Wang et al. 2017). The present study adds an important qualification to that literature: the energy benefit of occupancy-aware control depends on the credibility of the underlying occupancy measurement.

The findings also reinforce the value of moving beyond binary presence sensing. Traditional sensors such as PIR and CO_2 devices remain useful and are often simpler to deploy, but they do not always provide the room-level headcount information needed to estimate sensible gains, latent gains, and ventilation demand with sufficient fidelity (Chen et al. 2018; Rueda et al. 2020). In educational rooms, this distinction is important because two "occupied" states can imply very different HVAC requirements depending on whether there are five people in the room or twenty-five. From that perspective, the strongest result here is not merely that YOLOv7 can detect people in a classroom, but that, in the best configuration, it can produce a count schedule that is materially more relevant to HVAC than a static benchmark.

Several limitations should be acknowledged. First, the training images were generic rather than site-specific, which means the models were not explicitly tuned to the visual characteristics of room B5. Second, the empirical analysis covered one room and one 60-minute seminar, so the results should be read as a focused case study rather than as a general performance claim across all teaching spaces. Third, counts were sampled at one-minute intervals rather than extracted continuously from every frame, which is appropriate for schedule reconstruction but does not capture shorter transients. Fourth, the study reconstructed the control-relevant schedule and compared its simulated implications, but it did not implement a true closed-loop building management system. The load comparison is therefore best understood as a strong indication of control potential rather than as a direct measurement of realized energy savings. Fifth, the single room and single session mean that seasonal variation in daylighting and external climate was not captured; lighting and weather conditions can affect both detection accuracy and HVAC demand, so a cross-comparative study across rooms with different spatial configurations, occupant densities and seasons would be needed to establish generalisability and scalability beyond this case. Sixth, occlusion-induced under-counting is a known weakness of single-camera detection and was only partially mitigated here by comparing two viewpoints; multi-camera fusion and temporal tracking across frames are promising routes to more robust counts in high-density seating and are not implemented in the present study. Finally, because the EnergyPlus outputs are reported as normalised indices rather than as metered consumption, the energy implications should be read as indicative of relative potential rather than as measured operational savings; quantifying real savings will require instrumented monitoring, validation against indoor-air-quality sensors, or a metered closed-loop trial.

There are also practical deployment questions beyond counting accuracy. Vision-based systems raise privacy and governance issues that should be addressed explicitly in any real implementation. For operational building control, the most appropriate approach would likely be edge-based inference with storage minimization, limited retention, and the transmission of aggregate counts rather than identifiable video wherever possible. Future work should also test higher camera mounting positions, multi-view fusion, longer monitoring periods, and site-specific fine-tuning. Additional features such as activity level, equipment use, and window state would further improve the connection between detected occupancy and true internal gains, especially in mixed-mode or naturally ventilated spaces. Because the system observes people in an educational space, a real deployment should be governed by an explicit ethics and privacy protocol: institutional ethics approval and data-protection compliance (for example, GDPR), transparent signage and notification of those present, a defined lawful basis and retention schedule, on-device processing that discards raw frames once counts are extracted, and storage of only aggregate, non-identifiable counts. A further equity consideration is that occupancy-driven conditioning should preserve thermal equity and digital inclusion: the control logic must avoid systematically under-serving rooms or individuals that the camera tends to under-detect, such as occupants who are frequently occluded or seated at the frame edge, since biased sensing could otherwise translate into uneven comfort and ventilation provision across a building's users.

5. CONCLUSION

This study examined whether vision-based occupancy detection can provide a more useful HVAC input than a conventional static occupancy assumption in a university teaching space. Two YOLOv7 detectors were trained on the same annotated dataset but through different training pipelines and were deployed to two simultaneous recordings of the same seminar room from opposite corners. The extended cloud-GPU training configuration (Google Colab) outperformed the constrained training configuration (Roboflow) both on held-out validation and, more importantly, in real-room counting. The best-performing configuration, Google Colab with the front-right view, achieved MAE 0.75 persons, RMSE 1.09 persons, bias +0.07 persons, and normalized count accuracy 95.8%, indicating that the recovered schedule was both accurate and nearly unbiased.

When translated into HVAC-relevant terms, that same configuration reproduced the manually derived occupant gain with near-exact agreement, yielding 2.37 kW versus 2.36 kW from the front-right ground-truth schedule. Relative to the static benchmark of 4.48 kW, the best dynamic schedule was 47.2% lower. In the normalized EnergyPlus comparison, it corresponded to a winter-heating index of 73.3 and a summer-cooling index of 32.4, equivalent to reductions of 26.7% and 67.6%, respectively, against the static case. These values do not imply that every lower-looking detector output is beneficial; rather, they show that a validated dynamic schedule can materially change the inferred load picture of a teaching room when compared with a fixed benchmark.

The study therefore supports two conclusions. First, vision-based occupancy counting can provide a substantially better basis for teaching-space HVAC scheduling than static occupancy assumptions, provided that the detector and the camera view are carefully validated. Second, viewpoint and occlusion are not secondary implementation details but core determinants of whether a dynamic schedule is trustworthy enough for control use. A detector that systematically misses occupants may appear energy-efficient while simply under-representing the real room load. Future work should extend the approach to longer monitoring periods, multi-view or higher-mounted sensing, site-specific training data, and direct supervisory control integration so that the benefits of dynamic occupancy sensing can be tested under real operational conditions.

AUTHOR CONTRIBUTIONS

Rufus Dwamena: Conceptualisation, Data curation, Software, Investigation, Formal analysis, Writing – original draft, read and approved the final manuscript. **Wuxia Zhang**: Methodology, Software, Formal analysis, Writing – review and editing, read and approved the final manuscript. **Paige Tien**: Methodology, Validation, Writing – review and editing, read and approved the final manuscript. **John Calautit**: Supervision, Writing – original draft, Resources, read and approved the final manuscript. **Hao Sun**: Supervision, Writing – review and editing, Project administration.

COMPETING INTERESTS

The author declares no competing interests.

DATA ACCESSIBILITY

The video recordings analysed in this study contain identifiable individuals in a teaching space and therefore cannot be made publicly available for privacy and ethical reasons. The derived, non-identifiable data that support the findings of this study, namely the per-minute occupant counts, the reconstructed occupancy schedules, and the EnergyPlus model inputs and outputs, are available from the corresponding author on reasonable request.

REFERENCES

ASHRAE, 2020. ANSI/ASHRAE Standard 55-2020: Thermal Environmental Conditions for Human Occupancy. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

Bortolini, R, Rodrigues, R, Alavi, H, Vecchia, L F D and Forcada, N., 2022. Digital twins' applications for building energy efficiency: A review. Energies, 15(19): 7002. DOI: doi.org/10.3390/en15197002.

Chen, Z, Jiang, C and Xie, L., 2018. Building occupancy estimation and detection: A review. Energy and Buildings, 169: 260-270. DOI: doi.org/10.1016/j.enbuild.2018.03.084.

CIBSE, 2021. Guide A: Environmental design (2015, updated 2021). Available at https://www.cibse.org/knowledge-research/knowledge-portal/guide-a-environmental-design-2015 [Last accessed 18 April 2026].

Naylor, S, Gillott, M and Lau, T., 2018. A review of occupant-centric building control strategies to reduce building energy use. Renewable and Sustainable Energy Reviews, 96: 1-10. DOI: doi.org/10.1016/j.rser.2018.07.019.

Pang, Z, O'Neill, Z, Chen, Y, Zhang, J, Cheng, H and Dong, B., 2023. Adopting occupancy-based HVAC controls in commercial building energy codes: Analysis of cost-effectiveness and decarbonization potential. Applied Energy, 349: 121594. DOI: doi.org/10.1016/j.apenergy.2023.121594.

Pincott, J, Tien, P W, Wei, S and Calautit, J K., 2022. Indoor fire detection utilizing computer vision-based strategies. Journal of Building Engineering, 61: 105154. DOI: doi.org/10.1016/j.jobe.2022.105154.

Rueda, L, Agbossou, K, Cardenas, A, Henao, N and Kelouwani, S, 2020. A comprehensive review of approaches to building occupancy detection. Building and Environment, 180: 106966. DOI: doi.org/10.1016/j.buildenv.2020.106966.

Tien, P W, Wei, S, Calautit, J K, Darkwa, J and Wood, C., 2020. 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. DOI: doi.org/10.1016/j.enbuild.2020.110386.

Tien, P W, Wei, S and Calautit, J K., 2021a. A computer vision-based occupancy and equipment usage detection approach for reducing building energy demand. Energies, 14(1): 156. DOI: doi.org/10.3390/en14010156.

Tien, P W, Wei, S, Liu, T, Calautit, J, Darkwa, J and Wood, C., 2021b. 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. DOI: doi.org/10.1016/j.renene.2021.05.155.

Tien, P W, Wei, S, Calautit, J K, Darkwa, J and Wood, C., 2021c. Occupancy heat gain detection and prediction using deep learning approach for reducing building energy demand. Journal of Sustainable Development of Energy, Water and Environment Systems, 9(3): 1-31. DOI: doi.org/10.13044/j.sdewes.d8.0378.

Tien, P W, Wei, S, Darkwa, J, Wood, C and Calautit, J K., 2022a. Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality - a review. Energy and AI, 10: 100198. DOI: doi.org/10.1016/j.egyai.2022.100198.

Tien, P W, Wei, S, Calautit, J K, Darkwa, J and Wood, C., 2022b. 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. DOI: doi.org/10.1016/j.apenergy.2021.118336.

UK Government, 2021. Approved Document F, Volume 2: Buildings other than dwellings, 2021 edition, 15 December 2021. Available at https://www.gov.uk/government/publications/ventilation-approved-document-f [Last accessed 18 April 2026].

UNEP and GlobalABC, 2025. Global Status Report for Buildings and Construction 2024/2025. Nairobi: United Nations Environment Programme.

U.S. Department of Energy, 2026. EnergyPlus, 31 March 2026. Available at https://energyplus.net/ [Last accessed 18 April 2026].

Wang, C-Y, Bochkovskiy, A and Liao, H-YM., 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 17-24 June 2023, pp. 7464-7475. DOI: doi.org/10.1109/CVPR52729.2023.00721.

Wang, F, Feng, Q, Chen, Z, Zhao, Q, Cheng, Z, Zou, J, Zhang, Y, Mai, J, Li, Y and Reeve, H., 2017. Predictive control of indoor environment using occupant number detected by video data and CO_2 concentration. Energy and Buildings, 145: 155-162. DOI: doi.org/10.1016/j.enbuild.2017.04.014.

Wang, Z, Calautit, J, Wei, S, Tien, P W and Xia, L., 2022. Real-time building heat gains prediction and optimization of HVAC setpoint: An integrated framework. Journal of Building Engineering, 49: 104103. DOI: doi.org/10.1016/j.jobe.2022.104103.

Wang, Z, Calautit, J, Tien, P W, Wei, S, Zhang, W, Wu, Y and Xia, L., 2023b. An occupant-centric control strategy for indoor thermal comfort, air quality and energy management. Energy and Buildings, 285: 112899. DOI: doi.org/10.1016/j.enbuild.2023.112899.

Wei, S, Tien, P W, Calautit, J K, Wu, Y and Boukhanouf, R., 2020. Vision-based detection and prediction of equipment heat gains in commercial office buildings using a deep learning method. Applied Energy, 277: 115506. DOI: doi.org/10.1016/j.apenergy.2020.115506.

Wei, S, Tien, P W, Wu, Y and Calautit, J K., 2022a. A coupled deep learning-based internal heat gains detection and prediction method for energy-efficient office building operation. Journal of Building Engineering, 47: 103778. DOI: doi.org/10.1016/j.jobe.2021.103778.

Wei, S, Tien, P W, Chow, T W, Wu, Y and Calautit, J K., 2022b. Deep learning and computer vision based occupancy CO2 level prediction for demand-controlled ventilation (DCV). Journal of Building Engineering, 56: 104715. DOI: doi.org/10.1016/j.jobe.2022.104715.

Zhang, W, Calautit, J, Tien, P W, Wu, Y and Wei, S, 2024. Deep learning models for vision-based occupancy detection in high occupancy buildings. Journal of Building Engineering, 98: 111355. DOI: doi.org/10.1016/j.jobe.2024.111355.