Journal of
Engineering Sciences
Faculty of Engineering
Assiut University
JES
مجلة
العلوم الهندسية
كلية الهندسة
جامعة أسيوط
journal homepage: http://jesaun.journals.ekb.eg

General International Congress of Engineering (ICE)
9th International Architectural Conference of Assiut University (IACA-9)
Planning and Design for Tomorrow: Challenges, Experiences and Solutions
October 28-30, 2025, Hurghada, Egypt

Computer Vision-Based Occupancy Detection Systems and Approaches: A Review

Received July 2025; Revised August 2025; Accepted September 2025

Aya Shaheen[1]*
Yasser Mansour[2]
Hanan Sabry[3]
Fatma Fathy[4]
Hussein Hamza[5]

Keywords
Building Occupancy, Computer Vision, Energy Efficiency, HVAC Control, Deep Learning, Real-Time Monitoring, Thermal Comfort.

Abstract
Occupants are the core of the built environment. Developments in information technology facilitate gathering data on occupant presence and behaviour and learning from it through data-driven methods. Recent studies highlight the effectiveness of integrating occupancy data into HVAC control systems to reduce energy consumption while maintaining occupants' thermal comfort. This paper presents a comprehensive review of computer vision-based occupancy detection systems and approaches, mainly applied in the domain of controlling thermal environment. A systematic literature review methodology was employed to provide a structured process for the selection and analysis of studies. The paper aims to define occupancy parameters critical to occupant-centric HVAC control and personalized thermal comfort. It explores how computer vision techniques detect these parameters, with a focus on tools, deep learning models, dataset sizes, and evaluation metrics. Moreover, it summarizes various approaches that can be implemented for real-time indoor occupancy detection. The review analyses 43 papers published over the past decade. Findings emphasize the role of machine learning and computer vision in developing intelligent, occupant-centric control systems that optimize both comfort and energy efficiency. Finally, current challenges and future directions are discussed.

[1] aya.abdelhai@eng.asu.edu.eg -Assistant Lecturer, Dept. of Arch. Eng., Ain Shams University
[2] yasser_mansour@eng.asu.edu.eg - Professor, Dept. of Arch. Eng., Ain Shams University
[3] hanan_sabry@eng.asu.edu.eg - Professor, Dept. of Arch. Eng., Ain Shams University
[4] fatma.fathy@eng.asu.edu.eg -Assistant Professor, Dept. of Arch. Eng., Ain Shams University
[5] hussein.a.faried@eng.asu.edu.eg -Assistant Professor, Dept. of Arch. Eng., Ain Shams University

* Corresponding author

List of abbreviations

AI	Artificial Intelligence	OCC	Occupant-Centric Control
AP	Average Precision	PIR	Pyroelectric Infrared
CNN	Convolutional Neural Networks	PMV	Predicted Mean Vote
CV	Computer vision	PPD	Predicted Percentage of Dissatisfied
DNN	Deep Neural Networks	R- CLO	Real-time Clothing Insulation
FN	False Negatives	R-CNN	Region-based Convolutional Neural Networks
FP	False Positives	RGB	Red-Green-Blue
HVAC	Heating, Ventilation and air Conditioning	SSD	Single Shot Detector
IAQ	Indoor air quality	SVM	Support Vector Machine
IoT	Internet of Things	TN	True Negatives
LSTM	Long Short-Term Memory	TP	True Positives
ML	Machine Learning	YOLO	You Only Look Once

1. Introduction

Given that buildings contribute to approximately 40% of global energy use [1], recent studies highlight the significance of incorporating the human dimension in energy management. Factors as occupant count, behavior (including activity), thermal comfort preferences, and human interactions with the indoor environment significantly influence building energy performance. Therefore, by detecting real time data on occupants' presence, number, and behavior, building management systems can dynamically regulate HVAC systems. This adaptive control strategy ensures that indoor thermal conditions are continuously aligned with actual occupancy demand, thereby enhancing thermal comfort and reducing energy waste resulting from fixed schedules control.

Occupancy detection focuses on using sensors and monitoring equipment to capture the real-time occupancy information within a building [2], and often termed as "occupancy sensing" [3],[4]. Occupancy information can be classified into multiple levels: binary presence (occupied/vacant), occupancy count, identity, location, tracking and behavior within a built environment [5], [6]. Low-resolution data such as binary occupancy is sufficient for basic ventilation control, while higher-resolution information—such as occupant count and activity—enables advanced applications like occupant-centric HVAC control and personalized thermal comfort [7],[8]. The selection of the appropriate occupancy resolution depends on the system's objectives, while the selection of the sensing tool depends on occupancy resolution (the required parameters for occupancy detection), the building type, and privacy constraints. Therefore, managing data resolution is vital for deploying efficient and scalable occupancy-based building management systems.

Among the various techniques for acquiring occupancy information, computer vision is considered a promising non-intrusive method. Unlike wearable devices that require user participation or environmental sensors that often provide rough or delayed estimation, vision-based systems can unobtrusively monitor indoor environments using existing or low-cost camera infrastructure. Cameras function like the human eye; they can immediately recognize changes in occupants, even if they do not move, unlike PIR sensors that require people movement for counting them [9]. Computer vision can accurately detect and count occupants, recognize activities, and analyze posture or behaviors related to thermal comfort in real time, due to advancements in image processing and deep learning. This makes computer vision a scalable and flexible solution for occupancy-driven control in smart buildings, particularly for balancing between occupant comfort and energy efficiency.

This paper aims to define occupancy parameters, studied in the domain of occupant-centric HVAC control and personalized thermal comfort, and to explore how computer vision techniques are

employed to detect these parameters, focusing on tools, deep learning models used, dataset sizes, and model evaluation metrics. The paper starts with an overview of computer vision technology. Then, a comprehensive review of relevant studies is presented. The following sections highlight applications of computer vision in that domain, approaches adopted, image types, dataset characteristics, algorithms, and evaluation metrics. Finally, current research challenges and potential future directions are outlined, followed by the conclusion.

To guide this review, the following research questions are addressed:

RQ1: What occupancy parameters are most critical for occupant-centric HVAC control and personalized thermal comfort?

RQ2: How do computer vision techniques detect these occupancy parameters in indoor environments?

RQ3: What tools, deep learning models, dataset sizes, and evaluation metrics are commonly utilized in computer vision-based occupancy detection systems?

2. Research Method

A systematic literature review was conducted to gather information related to computer vision approaches for building occupancy sensing and detection. The search was conducted using Scopus, Web of Science and Google Scholar as search engines. Various keywords were used to collect the papers “computer vision”, “occupancy detection”, “occupancy monitoring”, “building”, “energy”, “thermal comfort”, “HVAC control”. The inclusion criteria were open access peer-reviewed articles published between 2016 and 2025 in English, focusing on the domain of occupant-centric HVAC control and personalized thermal comfort. Studies that introduce other sensors than computer vision systems were excluded. Based on these criteria, 43 papers published in journals and conferences were selected for review and analysis.

A qualitative synthesis was conducted by extracting and summarizing the following information from each study, including target occupancy parameters, data collection tool and type (camera and image type), computer vision (CV) approach and deep learning (DL) model used, Dataset characteristics (custom/open dataset, size), Evaluation metrics (e.g., accuracy, precision, recall, F1-score).

Following data extraction, studies were grouped into themes that emerged during the review process. These themes reflect both technical dimensions and application contexts and were used to structure the main body of the paper.

- Occupancy parameters
- Computer vision approaches
- Data collection and Dataset generation
- Deep learning models and algorithms
- Model Performance: Evaluation metrics

3. Computer Vision in Occupancy Detection: Literature Review

Computer vision is an essential application of artificial intelligence. It is an interdisciplinary field that deals with how computers can gain high-level understanding from digital images or videos. It can use computers and cameras to replace the human eye for the target object recognition, tracking, measurement, and for other visual problems. And then deal with the graphics so that the computer can achieve image processing capabilities even beyond the eye [10].

Traditional methods in computer vision have employed feature-based approaches such as scale invariant feature transform (SIFT), speeded up robust features (SURF), features from accelerated segment test (FAST), Hough transforms and geometric hashing, often integrated with machine

learning classifiers. Although these approaches remain relevant, the emergence of deep learning has a significant impact on that field. It offers greater accuracy and facilitates performing complex tasks [11]. The most dominant tool used in the processing stage of computer vision systems is CNN. CNNs are biologically inspired networks which are widely used for image recognition, classification, object detection, localization [11], and Semantic segmentation [10].

The integration of computer vision into building systems has gained significant attention for its potential to enhance occupant comfort and save energy. In this context, several studies have explored computer vision-based approaches to support adaptive HVAC control and personalized thermal comfort. This section presents a summary for the reviewed articles that applied computer vision for detecting occupancy, enabling occupant-centric HVAC control or personalized thermal comfort. The topics included during the review represents the phases of occupancy detection process based on CV Deep learning approach, as shown in Figure 1, starting from defining parameters and selecting suitable approach according to study aim, collecting data, preparing dataset, selecting suitable DL model/algorithm, training then evaluating the model.

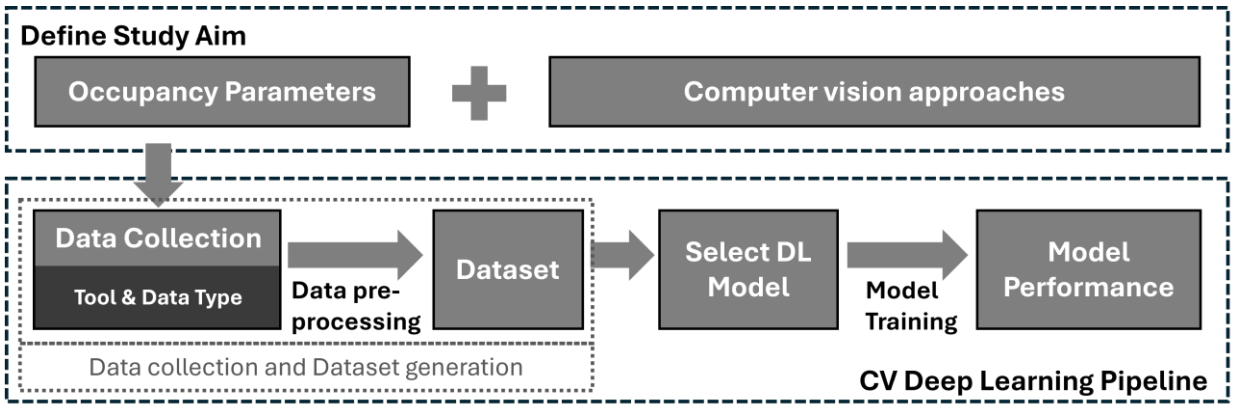

Figure 1: Occupancy Detection Process based on Computer Vision Deep Learning Approach

The 43 reviewed articles, presented in Table 1, are ordered chronologically, from recent to earlier publications. The table includes occupancy parameters detected in each study, image/video type, deep learning models/ algorithms, dataset sizes, and the metrics used for measuring performance of the employed model.

Table 1: Research Work on Computer Vision Approaches for Occupancy Detection

Ref.	year	Parameter				Data type	Dataset size	Model/ Algorithm	Performance	
		Count	Activity	Skin Temp.	Clothing				Metrics	Value (%)
[12]	2025		x		x	RGB images	Large dataset**	multi-task model (semi-supervised learning)	mAP	89.4*
[13]			x			RGB images	171	YOLOv7, Faster R-CNN	F1 scores	94*
[14]		x	x			RGB images	1200	Crowd-YOLO	Accuracy	90.96

[15]				x		Thermal images	5579	YOLOv11	mAP	94
Ref.	**year**	**Parameter**				**Data type**	**Dataset size**	**Model/ Algorithm**	**Performance**	
		Count	**Activity**	**Skin Temp.**	**Clothing**				**Metrics**	**Value (%)**
[16]	2024			x		Thermal images	Large dataset**	CNN	Precision Recall F1 Score Accuracy	99.44 99.26 99.32 99.51
[17]			x			RGB images	2359 images	HPE-based activity classification, DNN and YOLOv5	Accuracy	89
[18]			x		x	RGB images	Large dataset**	Proposed single-stage detection framework	Precision Recall F1 Score mAP@0.5	60.7 57.7 58.6 45.0
[19]			x		x	RGB images	6664 images	YOLOv5 model	Accuracy F1 score	95 93
[20]					x	RGB images	500 images	YOLO & GoogleNet	Accuracy	94.5*
[21]		x				RGB-Fisheye images	Large dataset**	CNN & (RAPiD)	Accuracy	98*
[22]		x				RGB videos	24 videos	PP-YOLOE	Accuracy	98.1
[23]		x				RGB images	377 images	SSD, MobileNetV2, Faster R-CNN InceptionV2 & YOLO	Accuracy	-
[24]	2023				x	RGB images	3 Large datasets**	YOLOv7	AP	82.4
			x			RGB videos		SlowFast network	Accuracy	93.9*
				x		RGB & Thermal images		OpenPifPaf	AP	71.9
[25]				x		RGB & IR images	-	CNN, Facial Landmark Detection algorithm	Accuracy	86
[26]			x			RGB images	-	YOLOv5	Precision Recall mAP	99.3 98 98.3
[27]				x		Thermal images	-	MediaPipe	-	-
[28]		x				RGB images	-	AS-DA-Net involving 2 CNN models	Average accuracy	89
[29]		x				RGB images	Large dataset**	YOLOv5	NRMSD *(real number)*	0.053

[30]	2022		x		x	RGB images	3000 images	YOLO	Accuracy	97
Ref.	**year**	**Parameter**				**Data type**	**Dataset size**	**Model/ Algorithm**	**Performance**	
		Count	**Activity**	**Skin Temp.**	**Clothing**				**Metrics**	**Value (%)**
[31]	2022		x	x	x	Thermal images (videos)	2422 short videos	CNN	Average accuracy	95.17
[32]			x		x	Thermal images	5263 images	YOLOv5 & DeepSort tracker	mAP50, Accuracy of Tracking	89.1, 99.5
[33]					x	RGB images	5400 images	YOLO	mAP	96
[34]		x				RGB videos	575 images	CNN	Accuracy	95.67
[35]		x				RGB Videos	Large dataset**	YOLOv3	-	-
[36]		x				RGB Videos	4405 images	YOLOX and new framework	Accuracy	99.2 *
[37]		x	x			RGB images	1500 images	Faster R-CNN with Inception V2	Accuracy	92.66*
[38]	2021		x			RGB images	1200 images	Faster R-CNN with Inception V2	Average accuracy	98.65
[39]			x			RGB images	870 images	DNN	PCP	92*
[40]			x			RGB images	480 images	Faster R-CNN	Accuracy	97.32*
[41]					x	RGB images	1000 images	CNN	Accuracy	94
[9]		x				RGB images	Large dataset**	YOLOv5	NRMSE *(real number)*	0.044 0.092
[42]		x				Thermal images	2346 images	CNN	Accuracy	94.1
[43]		x				RGB videos	-	YOLOv4	Accuracy	92.5
[44]	2020		x			RGB images	3472 images	CNN	Average accuracy	80.62
[4]		x	x			RGB videos	Large dataset**	YOLOv3	Accuracy	84
[45]			x			RGB videos	Large dataset**	ResNet-50 & LSTM network	Accuracy	99
[46]		x				RGB images	Large dataset**	CNN	Accuracy	70
[47]	2019		x			Depth images	Large dataset**	AlexNet	-	-
[48]			x		x	RGB images	300 images	OpenPose & (SVM) classifier	AP	94.2

[49]		x	x			RGB & depth images	-	CNN & (SVM) classifier	Accuracy	97.2 *
Ref.	**year**	**Parameter**				**Data type**	**Dataset size**	**Model/ Algorithm**	**Performance**	
		Count	**Activity**	**Skin Temp.**	**Clothing**				**Metrics**	**Value (%)**
[50]	2018				x	RGB & Thermal images	Large dataset**	GoogLeNet	Accuracy	90
[51]	2017	x				RGB video	-	SVM, CNN, K-means	Accuracy	95.3
[52]	2016	x				RGB video	Large dataset**	CNN, Adaboost Algorithm	Recall Precision	73 80

** Values presented in table 1 is the maximum accuracy (among classes) the model reached.*
***Large dataset refer to one of public open-source datasets*

To better understand the state of the art, the reviewed studies are discussed in the following sections in terms of their application contexts, tools, and models implementation.

4. Occupancy Parameters

Occupant count, behavior, comfort preferences, and interactions with the indoor environment—such as thermostat adjustments or window usage—significantly influence indoor thermal conditions and building energy performance. Computer vision approach is used to detect occupants' data in real time and integrate those data into control systems.

In HVAC control, computer vision-based systems help regulate heating, ventilation, and air conditioning by detecting real-time occupancy count. Many researchers [22] ,[21], [35], [36] ,[9],[52] worked on detecting occupancy count highlighting the potential of this approach on energy savings [40],[51],[42]. B. Yang et al. considered spatial distribution in addition to occupancy count, allowing zone-based conditioning and load prediction [29]. Meng et al. detected occupancy count for occupancy load estimation and air-conditioning predictive control [46]. However, research articles as [38], [44] recognized occupancy activity in indoor space providing data to form the real-time occupancy heat emission profiles known as the Deep Learning Influenced Profile (DLIP). Moreover, the potential of this method to save energy is studied [38]. Studies as [14], [37] , [43] managed to detect occupancy count/ density to achieve energy efficient ventilation control.

In thermal comfort management, computer vision techniques are used to assess activity level and posture, which are indicators of metabolic rate and individual thermal needs [31]. Yun et al., and Na et al. managed to detect occupants' activities to develop a metabolic rate prediction model, while Z. Wei et al. and H. Choi, Na, et al. classified clothing ensembles for calculating clothing insulation in real time. Many studies [12], [18], [19], [24], [31], [48] detected both activity and clothing for measuring Predicted Mean Vote (PMV) in real time. E. J. Choi et al. developed a model framework that classified five clothing ensembles in real-time, producing an accuracy of 86% in a real environment. In addition, they demonstrated that a PMV-based control system based on the model improved the thermal comfort of the occupants [33]. Researchers [24] ,[25], [31] detected skin temperature using computer vision. Zakka et al. developed a personal comfort model to automatically learn and classify different features from thermographic images to predict the thermal sensation of a single occupant [16]. Samal & Lone worked on a non-invasive, real-time temperature monitoring system designed for densely populated environments [15].

Each of these applications depends on varying levels of occupancy resolution. This review identified four parameters commonly investigated in the field of occupant-centric HVAC control and personalized thermal comfort: occupant count, activity recognition, clothes classification, and skin temperature detection.

5. Computer Vision Approaches

This section presents different computer vision approaches applied for detecting occupants' count, activity recognition, clothes classification, and skin temperature detection.

5.1. Occupancy Count

Computer Vision is considered a promising method for occupancy count, as it overcomes the limitations of traditional PIR sensors (especially low accuracy in low-level motion) and CO2 sensors. Most studies [14] ,[29] , [9], [43], [4] used object detection YOLO model. It works on finding the location of an object in an image and classifying the object type. Subsequently, an algorithm is added for counting the virtual boxes (i.e., bounding boxes) assigned to occupants detected by YOLO. If images are captured in real time, the number of occupants can also be counted in real time by counting the number of bounding boxes in each image [9]. This is called scene-based counting method. In this method, the camera should be facing indoors—thus, it is subjected to privacy issues. Furthermore, it is suitable for relatively small rooms because all occupants must fit within the camera's field of view. For scene-based counting, most researchers used RGB images. However, Kraft et al. used low-quality thermal images to cope with privacy issues [42].

Line-based count is a method of counting the number of people entering and leaving an imaginary line and estimating the number of occupants in the room. This method is more complex than scene-based counting, as it requires not only identifying the individuals but also monitoring their movement patterns (occupancy tracking). Cameras are often installed on ceilings above the room entrance. Therefore, there were relatively few privacy issues. Occupancy counting is possible even when the room is large. However, if many people pass through the line simultaneously, there is a high probability of an error due to occlusion. Recently, Gursel Dino et al. proposed a hybrid method that combines line-based and scene-based counting [35]. They utilized deep learning architectures to estimate the number of people in large, crowded spaces using multiple cameras. Various vision techniques (head detection, background elimination, head tracking) are implemented in three methods: (i) a method that instantaneously counts people in a scene, (ii) a method that incrementally counts people entering/exiting a room and (iii) a combination of the first two methods. These methods were applied in a classroom with heavy occlusions and resulted in a high prediction capacity when compared to ground truth measurements [35].

Sun et al. proposed a new occupancy counting method consists of five phases: (i) occupancy detection by generating bounding boxes around occupants, (ii) occupancy tracking, the tracker uses these bounding boxes to output the moving trajectories of the occupants, (iii) trajectory hypothesis, that refines the moving trajectories to improve accuracy, (iv) occupancy counting, and (v) multi-camera fusion [22]. Another paper [36] proposed a novel fusion framework for occupancy detection and estimation based on two different perspectives. First, they designed a head detection method combined with indoor scene knowledge to filter false positives and recover missed detection. Second, they proposed a two-vision entrance counting method to refine the predicted results. They observed that indoor-view method (cameras at the room interior) did not work when many people exist at the room entrance, while the overhead method (cameras at the room entrance) was prone to error when

many people concurrently enter and exit. Therefore, they combined the two-vision situations to refine the counting results [36].

5.2. Activity Recognition

Activity recognition has recently gained significant attention, due to advancements in computer vision technology. Computer vision works on extracting joint positions or color distributions of occupants from images and classifying these patterns as specific activities using predefined data features. Both depth and RGB images are commonly employed. In some studies, researchers have combined RGB and depth images for improved accuracy [49]. Moreover, Liu et al. [18] used an object detection model that identified both actions and clothing in thermal images.

Among the various human pose estimation networks, OpenPose is one of the most widely used. It can simultaneously detect human skeletons, recognize the joints of multiple individuals, and track skeleton key points from images. Classifying Activities based on joint coordinates is a well-established and widely adopted approach in this domain [19], [17]. However, other studies employed object detection models to recognize occupants' activities through color distributions [18], [44], [38]. Yun et al. proposed a new framework for enhancing detection and classification accuracy by incorporating a third phase - object detection - following joint recognition and activity classification [17]. This approach enabled the identification of objects associated with human activities, leading to higher detection precision. While some recent studies focused on recognizing individual occupant activities, others [45] targeted group-level activities such as presentations, meetings, or in activity.

5.3. Defining Clothing condition/ insulation

Clothing insulation refers to thermal resistance of clothes. It is one of the most important thermal comfort adjustments available to building occupants. It represents one of the two personal factors affecting the determination of the predicted mean vote (PMV) and predicted percentage of dissatisfied (PPD)-the most widely used thermal comfort indices.

Various studies work on defining real time clothing insulation and integrating it into occupant-centric HVAC control. Two approaches are presented in this domain. The first approach is the use of RGB images. Deep learning models trained to learn the shape of clothes from images, classifying the types of garments worn by the occupants, then calculating the corresponding clothing insulation value. Wei at al. categorized the garments into light, medium and heavy clothing [20]. Another study summarized the fiber and corresponding garment insulation for the 16 selected garments organized into five categories: top, bottom, outer, dress, and pajamas [33]. E. J. Choi et al. developed a model framework that classified five clothing ensembles in real-time, producing an accuracy of 86% in a real environment [33].

The second approach involves using thermal images allowing temperature detection from images. First, the clothing status together with a key body points detector locate the person's skin region and clothes region, allowing the measurement of skin temperature and clothes temperature [32]. Liu et al. used CNN model to recognize an occupant's clothes type that helps to differentiate the skin region from the clothing-covered region, consequently, calculating the skin temperature and the clothes temperature [31]. While this method is non-invasive, its accuracy can be affected by practical limitations, such as the distance between the infrared (IR) camera and the target, and the fact that only uncovered body parts can be accessed by the sensors [33].

5.4. Skin Temperature Estimation

Thermal cameras are used to measure skin temperature, by applying image-processing techniques that detect specific body parts (skin area). The recorded temperature for the detected area is then used

to infer thermal sensation, as skin temperature reflects an individual's thermal perception. Among the exposed body parts, the face has been most frequently analyzed, [25] , [15] as it provides a reliable indicator of thermal sensation and is easily accessible for measurement. Several studies have also focused on other areas, such as the hands [24], [27] or multiple parts of the upper body including hands, forearm, and head [27]. The machine learning library OpenPifPaf is employed for posture and body parts detection, providing detailed key point annotation for the face, hands, and feet to enable accurate pose estimation from RGB images [24]. By using synchronized RGB and IR image streaming, the model can identify the coordinates of the head and hands in the RGB images and extract the corresponding temperature data from the IR thermal images [24]. Samal & Lone introduced a system that employs YOLO models for face detection and utilizes a regression framework for mapping pixel to temperature values [15] . Lyu et al. used MediaPipe, an opensource machine learning framework and OpenCV library for face mesh generation. [27] while Intharachathorn et al. used CNN-Facial Landmark Detection algorithm [25].

6. Data Collection and Dataset Generation

Effective computer vision systems rely on the quality and diversity of input data. This section outlines the common image types used (e.g., RGB, infrared/thermal, depth) and discusses key practices in dataset collection, annotation, and preprocessing essential for training robust occupancy detection models.

6.1. Data Collection Tool and Data Type

The type of image data significantly influences the resolution of occupant information. Such data can be acquired from either still images or continuous video streams. Image modalities are commonly categorized into three types: RGB, depth, and thermal. RGB images capture color and detailed visual features. They are usually taken with standard RGB cameras, webcams are most widely used because they are affordable and widely available. Depth images show the distance between objects and the camera [49]. Initially, these images were captured using commercial 3D stereovision systems. However, recent studies have increasingly used more cost-effective devices like the Microsoft Kinect[47] and Intel RealSense depth cameras. Thermal images help ensure privacy and operate effectively under varying lighting conditions. Moreover, it reflects the temperature of objects. These can be obtained from high-end thermal cameras [15], [31], [27], infrared cameras [24], [16], [25], or recently, low-cost infrared array sensors [50]. These sensors have gained attention in recent research for their balance of performance and cost.

6.2. Dataset Preparation and Use

Dataset preparation is a fundamental step in developing and evaluating computer vision-based occupant detection systems. Images or videos typically undergo a preprocessing phase that includes labeling and data augmentation. Labeling is crucial for training, as it involves annotating the data to identify key occupancy features such as presence, count, location, or activity. Accurate and consistent labeling ensures that the model learns meaningful patterns, directly influencing detection performance. This process may be carried out manually using annotation tools or through automated techniques to create ground truth data for supervised learning.

Once the dataset is prepared, it is typically divided into three subsets: training, validation, and testing. Researchers may use either custom or open-source datasets. Custom datasets are often collected in specific indoor environments—such as offices, classrooms, or laboratories—to reflect unique architectural and occupancy conditions. Alternatively, data may be collected from the internet and

manually labeled to suit specific research objectives. The size of datasets vary widely across studies. H. Choi et al. [30] used 3000 images for training the model, while Zhou et al. prepared only 171 images in their dataset [13]. The largest number among the reviewed articles was 6664 images, used by E. J. Choi et al. for training the model [19]. Figure 2 shows the different sizes for custom datasets according to the reviewed articles.

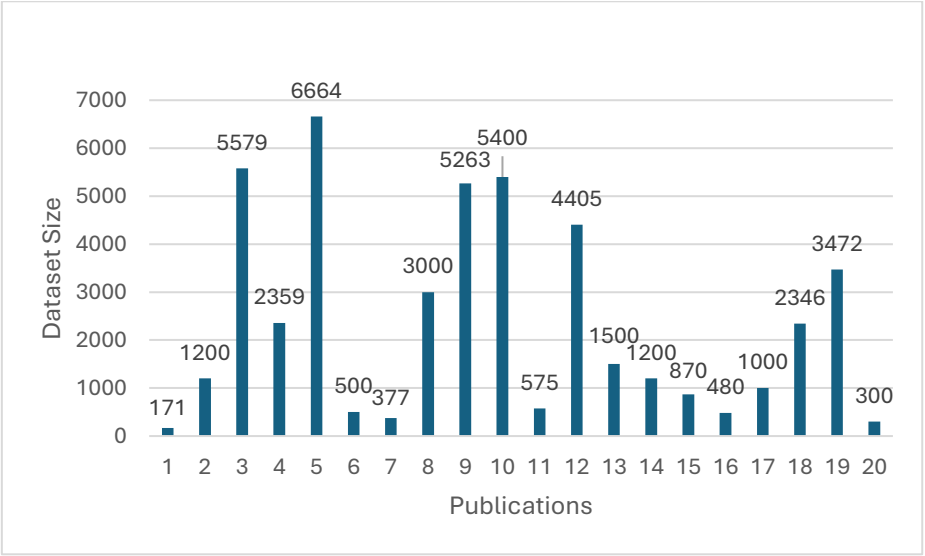

Figure 2: Size of Custom Datasets Used in Reviewed Articles

However, dataset generation is time-consuming, expensive and lack standardized formats, which makes it hard to compare between different research results. Thus, the decision to use custom or open datasets usually depends on application goals, privacy concerns, and the desired level of occupancy resolution. Several publicly available image and video datasets have been utilized in studies focused on occupancy counting, activity recognition, clothing detection, and related occupant parameters. These datasets vary in modality, environment, and annotation detail, supporting tasks such as person detection, pose estimation, and behaviour analysis. Common examples include COCO [9], [29], BRAINWASH [46], Wanda mall [46], Qinghai showroom [46] and SCUT-HEAD [35] for occupancy counting; Kinetics [24] and AMI Corpus [45] for activity recognition; DeepFashion [20], [24], Fashion-MNIST [20], IndoFashion [20] and ImageNet [50] for clothing classification. Researchers often select datasets based on the specific parameter of interest or combine multiple datasets to train and evaluate multi-task models. Numerous studies use offline datasets for both training and testing, failing to evaluate their methods in real-world settings, which can lead to unrealistic performance expectations [15]. As a result, there is an increasing demand for publicly available, large-scale, multi-modal datasets that reflect real-world indoor conditions. These datasets can support the development of general models that are scalable.

7. Deep Learning Models and Algorithms

Deep learning-based object detection models such as faster R-CNN [23], [37], [38], R-CNN [9], single shot detector (SSD) [23], and YOLO [9], [35], [4], [43], [29] are known for their high detection accuracy. Their performance is attributed to their training on very large image datasets, which enables them to effectively learn and recognize complex visual patterns. Many researchers compared different object detection models in terms of speed and accuracy. Alsultan & Mohammad stated that YOLOv7 and Faster R-CNN are of high accuracy, while SSD and RetinaNet is of moderate accuracy. They added that in terms of speed YOLOv7 and SSD are fast and suitable for real time applications [53]. Shen et al. used scaled YOLO model, after transfer learning, and proved high accuracy than other models (Faster R-CNN, SSD, YOLOv3) [54]. W. Zhang et al. compared different versions of

YOLO model with SSD and Faster R-CNN. They stated that YOLOv8n model has the least training time, while YOLOv8x model performed with the highest precision [23].

It was observed while reviewing the articles that YOLO model has been widely used for occupancy counting [9],[35],[4],[43],[29]. RAPiD model used for counting occupants in crowded scenes. It is suited for occupancy detection applications requiring precise person localization and robustness to occlusion [21]. For activity recognition and body part recognition, pre-trained CNN models and OpenPose, a joint tracking model for detecting human skeletons, are commonly employed [17],[19]. OpenPifPaf [24] is also used for human pose estimation. However, many studies used YOLO for activity recognition [26], [4] and classifying clothing ensembles [19], [30], [33]. YOLO uses a single-stage object detection network, which enables faster detection speed compared to other models. Consequently, YOLO is widely adopted in real-time applications.

Moreover, It was observed that on using pre-trained models, users can develop fast deep learning models that perform well with small custom datasets. A pre-trained model refers to a model already trained using a large dataset and well-designed CNN architecture. This process is called fine tuning or transfer learning. In the reviewed studies, pre-trained models such as Inception [23], [37], [38], GoogLeNet [20], AlexNet [47], and ResNet [45] were used.

8. Model Performance: Evaluation Metrics

The performance of occupancy detection systems is analyzed using various metrics such as accuracy, Root Mean Square Error (RMSE), precision, recall, and F1-score. Among these metrics, accuracy is one of the most used scores[5]. In general, the most commonly used metrics include precision, recall, and the F1-score, which are particularly important in classification-based tasks such as presence detection and activity recognition. Precision quantifies the proportion of true positives among all positive predictions, assessing the model's capability to avoid false positives. On the other hand, Recall calculates the proportion of true positives among all actual positives, measuring the model's ability to detect all instances of a class. The F1-score, the harmonic mean of precision and recall, offers a balanced metric in cases of class imbalance. For object detection and occupant counting tasks, metrics like mean Average Precision (mAP) and Intersection over Union (IoU) are widely adopted. Mean Average Precision (mAP) summarizes detection performance across multiple thresholds and classes, while IoU quantifies the overlap between predicted and ground truth bounding boxes. In some studies, especially those focusing on multi-occupant scenarios, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are used to evaluate counting accuracy. The choice of metric often depends on the specific task (e.g., classification, detection, counting) and the application domain, highlighting the need for careful selection to ensure meaningful and comparable results across studies. B. Yang et al. and H. Choi, Um, et al. used the normalized root-mean-square deviation (NRMSD) as a performance metric to assess accuracy [29], [9]. It represents the difference between the actual and predicted values, and the closer it is to zero, the closer the algorithm detection value is to the true value, and the more accurate the detection [39]. E. J. Choi et al. used mean squared error (MSE) and percentage of corrected parts (PCP) to quantify model accuracy [39].

In many cases, the comparison between models should go beyond common metrics. According to the developers of YOLO model, the speed of inference can be as critical as accuracy, especially in real-time object detection scenarios [55]. A recently published paper [56] stated that there are alternative metrics that should be taken into consideration such as computer hardware and software requirements, computation time, robustness to changing conditions, sensitivity to model and forecast uncertainty, data requirements, and implementation effort.

9. Discussion

Computer vision-based methods for occupancy detection face several limitations and challenges that hinder large-scale deployment, despite their growing effectiveness. One of the primary concerns is privacy, especially in workplaces and homes, where continuous visual monitoring may be perceived as intrusive. While alternatives like depth sensors [49] or thermal imaging [42] can mitigate privacy risks, they often come at a higher cost or lower accuracy. Future research can focus on developing privacy-preserving techniques, such as silhouette extraction which allow occupant detection without revealing personal identity [54],[57]. Another challenge arises in high-density occupancy scenarios, where image recognition using computer vision may suffer from occlusion and overlap [35], leading to detection errors that compromise the accuracy of HVAC control. This limitation has been noted in several studies, highlighting the need for future research to incorporate deep learning and 3D imaging techniques to enhance accuracy and evaluate the robustness of CV algorithms in more complex environments [8]. Moreover, Lighting variability and camera placement limitations [22] can also affect detection accuracy, especially in dynamic indoor environments with moving objects or multiple occupants. Another growing area is the integration of multi-modal sensor fusion, where CV data is combined with Wi-Fi, CO_2, or wearable sensors to improve robustness under occlusion, poor lighting, and high-density occupancy conditions [58].

Another key challenge is the computational demand and infrastructure requirements associated with real-time video processing, which may require edge computing capabilities or high-performance servers, increasing system complexity and energy use [56]. However, the emergence of lightweight deep learning models and edge AI frameworks facilitate real-time, energy-efficient processing on embedded systems, reducing dependency on cloud infrastructure. Research is also shifting toward improving model generalization through transfer learning and domain adaptation [59], addressing decrease in performance when models are deployed across different building types, cultural contexts, or environmental conditions. Finally, there is an urgent need for standardized, diverse, and annotated open-source datasets that represent realistic indoor scenarios. This will result in having objective criteria for benchmarking models [58]. These future directions are essential for developing scalable, occupant-centric, and energy- efficient smart building systems that respond adaptively to dynamic human behavior.

10. Conclusion

This review presented a comprehensive analysis of computer vision-based occupancy detection approaches on the domain of occupant-centric HVAC control and personalized thermal comfort, published between 2016 to 2025. A critical synthesis of recent peer-reviewed studies revealed a growing shift toward the use of high-resolution image modalities, including RGB, depth, and thermal imaging, each with varying trade-offs in accuracy, privacy, and environmental robustness. Furthermore, this review has examined the evolution of deep learning models, noting the dominance of convolutional neural networks (CNNs), YOLO-based detectors, and hybrid architectures in recent implementations. These models have enabled improved occupant detection, counting, and activity inference in real-world conditions. Evaluation metrics commonly used in the literature -such as precision, recall, F1-score, and mean average precision (mAP)- were discussed to assess model performance, with particular attention to challenges arising from multi-occupant scenes, lighting variation, and occlusion. Despite significant advancements, key challenges remain, including privacy concerns, dataset limitations, model generalizability, and real-time deployment constraints. Figure 3 summarizes the key findings of this review through a diagram that connects occupancy parameters

(e.g., count, activity, clothing, and skin temperature) with suitable computer vision techniques, data collection processes, model selection strategies, and performance evaluation criteria.

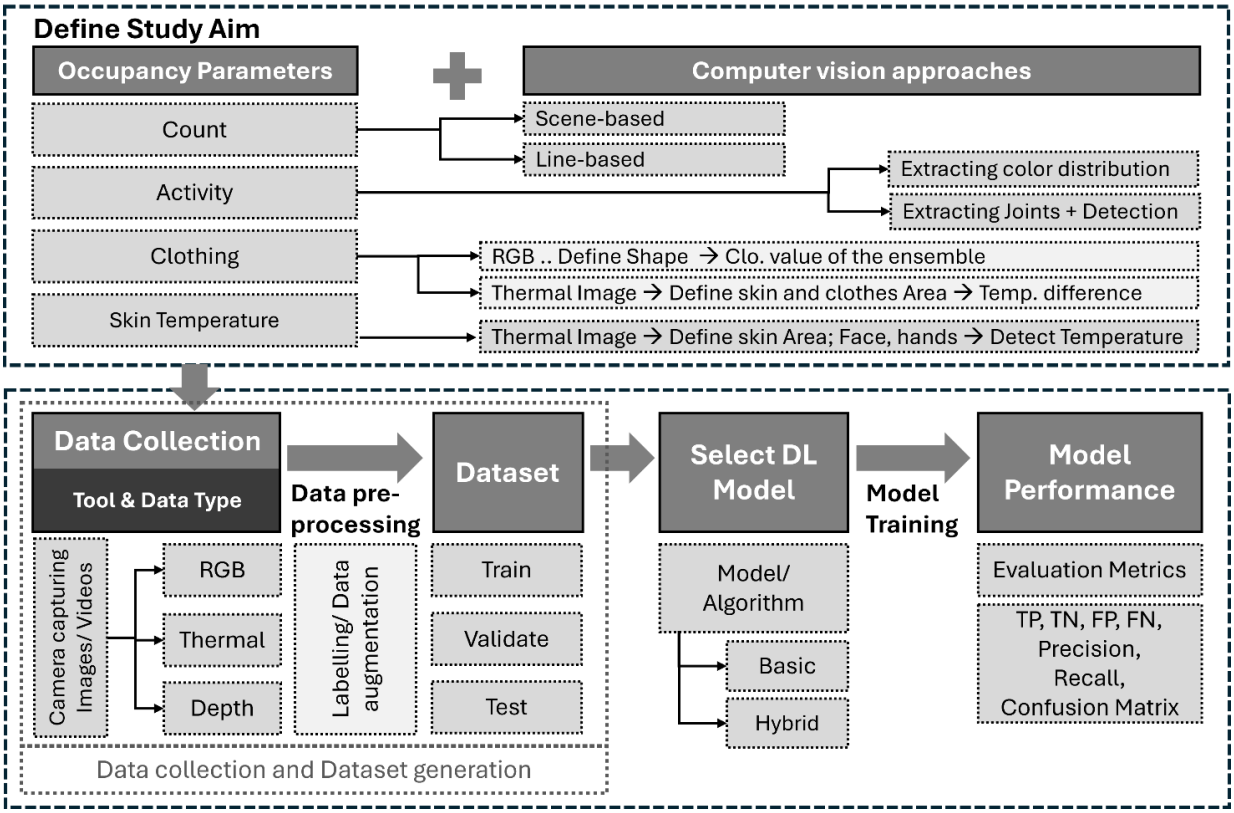

Figure 3: Overview of Key Findings in Computer Vision-Based Occupancy Detection

Future research must focus on developing privacy-aware, multimodal, and edge-optimized solutions, as well as on the standardization of benchmarking datasets and metrics. Finally, the integration of robust computer vision systems into building management frameworks holds great promise for enabling intelligent, occupant-aware environments that optimize both comfort and energy efficiency.

References

[1] K. Sun, Q. Zhao, and J. Zou, "A review of building occupancy measurement systems," Jun. 01, 2020, *Elsevier Ltd*. doi: 10.1016/j.enbuild.2020.109965.

[2] D. Trivedi and V. Badarla, "Occupancy detection systems for indoor environments: A survey of approaches and methods," Oct. 01, 2020, *SAGE Publications Ltd*. doi: 10.1177/1420326X19875621.

[3] H. Choi, C. Y. Um, K. Kang, H. Kim, and T. Kim, "Review of vision-based occupant information sensing systems for occupant-centric control," Oct. 01, 2021, *Elsevier Ltd*. doi: 10.1016/j.buildenv.2021.108064.

[4] I. Mutis, A. Ambekar, and V. Joshi, "Real-time space occupancy sensing and human motion analysis using deep learning for indoor air quality control," *Autom Constr*, vol. 116, p. 103237, 2020.

[5] L. Rueda, K. Agbossou, A. Cardenas, N. Henao, and S. Kelouwani, "A comprehensive review of approaches to building occupancy detection," Aug. 01, 2020, *Elsevier Ltd*. doi: 10.1016/j.buildenv.2020.106966.

[6] T. Li *et al.*, "A systematic review and comprehensive analysis of building occupancy prediction," Apr. 01, 2024, *Elsevier Ltd*. doi: 10.1016/j.rser.2024.114284.

[7] J. Chen, H. Chen, and X. Luo, "Collecting building occupancy data of high resolution based on WiFi and BLE network," *Autom Constr*, vol. 102, pp. 183–194, Jun. 2019, doi: 10.1016/j.autcon.2019.02.016.

[8] J. Wang, L. Jiang, H. Yu, Z. Feng, R. Castaño-Rosa, and S. jie Cao, "Computer vision to advance the sensing and control of built environment towards occupant-centric sustainable development: A

critical review," *Renewable and Sustainable Energy Reviews*, vol. 192, Mar. 2024, doi: 10.1016/j.rser.2023.114165.

[9] H. Choi, C. Y. Um, K. Kang, H. Kim, and T. Kim, "Application of vision-based occupancy counting method using deep learning and performance analysis," *Energy Build*, vol. 252, Dec. 2021, doi: 10.1016/j.enbuild.2021.111389.

[10] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," May 01, 2021, *Elsevier Ireland Ltd*. doi: 10.1016/j.cosrev.2021.100379.

[11] L. McMillan and L. Varga, "A review of the use of artificial intelligence methods in infrastructure systems," Nov. 01, 2022, *Elsevier Ltd*. doi: 10.1016/j.engappai.2022.105472.

[12] S. Jung, J. Jeoung, M. Kong, and T. Hong, "Occupant activities and clothes detection based on semi-supervised learning for occupant-centric thermal control," *Build Environ*, vol. 267, p. 112178, 2025.

[13] S. Zhou, W. Zhang, P. W. Tien, and J. Calautit, "Evaluating single-shot and two-stage vision-based detection of occupancy activity in energy-efficient buildings," *Energy Build*, vol. 345, Oct. 2025, doi: 10.1016/j.enbuild.2025.116017.

[14] N. Yue, L. Li, M. Caini, and X. Xie, "Occupant information computer vision sensing-based displacement ventilation in large space building for improving indoor environment and energy efficiency," *Build Environ*, vol. 269, Feb. 2025, doi: 10.1016/j.buildenv.2024.112364.

[15] A. Samal and H. R. Lone, "Thermal vision: Pioneering non-invasive temperature tracking in congested spaces," *Smart Health*, vol. 36, Jun. 2025, doi: 10.1016/j.smhl.2025.100576.

[16] V. G. Zakka, M. Lee, R. Zhang, L. Huang, S. Jung, and T. Hong, "Non-invasive vision-based personal comfort model using thermographic images and deep learning," *Autom Constr*, vol. 168, Dec. 2024, doi: 10.1016/j.autcon.2024.105811.

[17] J. Y. Yun, E. J. Choi, M. H. Chung, K. W. Bae, and J. W. Moon, "Performance evaluation of an occupant metabolic rate estimation algorithm using activity classification and object detection models," *Build Environ*, vol. 252, p. 111299, 2024.

[18] S. Jung, J. Jeoung, T. Hong, and H. Jang, "Vision-based multi-label detection framework for capturing occupant action and clothing information using large-scale dataset," *Build Environ*, vol. 257, Jun. 2024, doi: 10.1016/j.buildenv.2024.111537.

[19] E. J. Choi, J. Y. Yun, Y. J. Choi, M. C. Seo, and J. W. Moon, "Impact of thermal control by real-time PMV using estimated occupants personal factors of metabolic rate and clothing insulation," *Energy Build*, vol. 307, Mar. 2024, doi: 10.1016/j.enbuild.2024.113976.

[20] Z. Wei, J. K. Calautit, S. Wei, and P. W. Tien, "Real-time clothing insulation level classification based on model transfer learning and computer vision for PMV-based heating system optimization through piecewise linearization," *Build Environ*, vol. 253, Apr. 2024, doi: 10.1016/j.buildenv.2024.111277.

[21] J. Konrad, M. Cokbas, P. Ishwar, T. D. C. Little, and M. Gevelber, "High-accuracy people counting in large spaces using overhead fisheye cameras," *Energy Build*, vol. 307, p. 113936, 2024.

[22] K. Sun, X. Wang, T. Xing, S. Liu, and Q. Zhao, "High-accuracy occupancy counting at crowded entrances for smart buildings," *Energy Build*, vol. 319, Sep. 2024, doi: 10.1016/j.enbuild.2024.114509.

[23] W. Zhang, J. Calautit, P. W. Tien, Y. Wu, and S. Wei, "Deep learning models for vision-based occupancy detection in high occupancy buildings," *Journal of Building Engineering*, vol. 98, Dec. 2024, doi: 10.1016/j.jobe.2024.111355.

[24] M. Rida, M. Abdelfattah, A. Alahi, and D. Khovalyg, "Toward contactless human thermal monitoring: A framework for Machine Learning-based human thermo-physiology modeling augmented with computer vision," *Build Environ*, vol. 245, Nov. 2023, doi: 10.1016/j.buildenv.2023.110850.

[25] K. Intharachathorn, D. Jareemit, and S. Watcharapinchai, "Potential use of an extended-distance thermal imaging camera for the assessment of thermal comfort in multi-occupant spaces," *Build Environ*, vol. 246, Dec. 2023, doi: 10.1016/j.buildenv.2023.110949.

[26] K. Li, J. Li, Y. Zhou, L. Liu, Y. Zhu, and Z. Yu, "A Study on Multiple Occupant Behavior Detection Based on Computer Vision Technique," in *Building Simulation Conference Proceedings*,

International Building Performance Simulation Association, 2023, pp. 3888–3892. doi: 10.26868/25222708.2023.1736.

[27] J. Lyu, H. Du, Z. Zhao, Y. Shi, B. Wang, and Z. Lian, "Where should the thermal image sensor of a smart A/C look?-Occupant thermal sensation model based on thermal imaging data," *Build Environ*, vol. 239, Jul. 2023, doi: 10.1016/j.buildenv.2023.110405.

[28] Z. Cui, Y. Sun, D. Gao, J. Ji, and W. Zou, "Computer-vision-assisted subzone-level demand-controlled ventilation with fast occupancy adaptation for large open spaces towards balanced IAQ and energy performance," *Build Environ*, vol. 239, p. 110427, 2023.

[29] B. Yang, Y. Liu, P. Liu, F. Wang, X. Cheng, and Z. Lv, "A novel occupant-centric stratum ventilation system using computer vision: Occupant detection, thermal comfort, air quality, and energy savings," *Build Environ*, vol. 237, Jun. 2023, doi: 10.1016/j.buildenv.2023.110332.

[30] H. Choi, B. Jeong, J. Lee, H. Na, K. Kang, and T. Kim, "Deep-vision-based metabolic rate and clothing insulation estimation for occupant-centric control," *Build Environ*, vol. 221, Aug. 2022, doi: 10.1016/j.buildenv.2022.109345.

[31] J. Liu, I. W. Foged, and T. B. Moeslund, "Automatic estimation of clothing insulation rate and metabolic rate for dynamic thermal comfort assessment," *Pattern Analysis and Applications*, vol. 25, no. 3, pp. 619–634, Aug. 2022, doi: 10.1007/s10044-021-00961-5.

[32] J. Liu, I. W. Foged, and T. B. Moeslund, "Clothing Insulation Rate and Metabolic Rate Estimation for Individual Thermal Comfort Assessment in Real Life," *Sensors*, vol. 22, no. 2, Jan. 2022, doi: 10.3390/s22020619.

[33] E. J. Choi, B. R. Park, N. H. Kim, and J. W. Moon, "Effects of thermal comfort-driven control based on real-time clothing insulation estimated using an image-processing model," *Build Environ*, vol. 223, Sep. 2022, doi: 10.1016/j.buildenv.2022.109438.

[34] Y. Yang, Y. Yuan, T. Pan, X. Zang, and G. Liu, "A framework for occupancy prediction based on image information fusion and machine learning," *Build Environ*, vol. 207, Jan. 2022, doi: 10.1016/j.buildenv.2021.108524.

[35] I. Gursel Dino, E. Kalfaoglu, O. K. Iseri, B. Erdogan, S. Kalkan, and A. A. Alatan, "Vision-based estimation of the number of occupants using video cameras," *Advanced Engineering Informatics*, vol. 53, Aug. 2022, doi: 10.1016/j.aei.2022.101662.

[36] K. Sun, P. Liu, T. Xing, Q. Zhao, and X. Wang, "A fusion framework for vision-based indoor occupancy estimation," *Build Environ*, vol. 225, Nov. 2022, doi: 10.1016/j.buildenv.2022.109631.

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

[38] P. W. Tien, S. Wei, J. K. Calautit, J. Darkwa, and C. Wood, "Vision-based human activity recognition for reducing building energy demand," *Building Services Engineering Research and Technology*, vol. 42, no. 6, pp. 691–713, Nov. 2021, doi: 10.1177/01436244211026120.

[39] E. J. Choi, J. W. Moon, J. H. Han, and Y. Yoo, "Development of a deep neural network model for estimating joint location of occupant indoor activities for providing thermal comfort," *Energies (Basel)*, vol. 14, no. 3, Feb. 2021, doi: 10.3390/en14030696.

[40] P. W. Tien, S. Wei, and J. Calautit, "A computer vision-based occupancy and equipment usage detection approach for reducing building energy demand," *Energies (Basel)*, vol. 14, no. 1, Jan. 2021, doi: 10.3390/en14010156.

[41] H. Choi, H. S. Na, T. Kim, and T. Kim, "Vision-based estimation of clothing insulation for building control: A case study of residential buildings," *Build Environ*, vol. 202, p. 108036, Sep. 2021, doi: 10.1016/J.BUILDENV.2021.108036.

[42] M. Kraft, P. Aszkowski, D. Pieczyński, and M. Fularz, "Low-cost thermal camera-based counting occupancy meter facilitating energy saving in smart buildings," *Energies (Basel)*, vol. 14, no. 15, Aug. 2021, doi: 10.3390/en14154542.

[43] J. Wang, J. Huang, Z. Feng, S. J. Cao, and F. Haghighat, "Occupant-density-detection based energy efficient ventilation system: Prevention of infection transmission," *Energy Build*, vol. 240, Jun. 2021, doi: 10.1016/j.enbuild.2021.110883.

[44] P. W. Tien, S. Wei, J. K. Calautit, J. Darkwa, and C. Wood, "A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand-driven control solutions," *Energy Build*, vol. 226, Nov. 2020, doi: 10.1016/j.enbuild.2020.110386.

[45] G. A. Florea and R. C. Mihailescu, "Multimodal deep learning for group activity recognition in smart office environments," *Future Internet*, vol. 12, no. 8, Aug. 2020, doi: 10.3390/FI12080133.

[46] Y. bo Meng, T. yue Li, G. hui Liu, S. jun Xu, and T. Ji, "Real-time dynamic estimation of occupancy load and an air-conditioning predictive control method based on image information fusion," Apr. 15, 2020, *Elsevier Ltd*. doi: 10.1016/j.buildenv.2020.106741.

[47] H. Na, J.-H. Choi, H. Kim, and T. Kim, "Development of a human metabolic rate prediction model based on the use of Kinect-camera generated visual data-driven approaches," *Build Environ*, vol. 160, p. 106216, 2019.

[48] M. Zang, Z. Xing, and Y. Tan, "IoT-based personal thermal comfort control for livable environment," *Int J Distrib Sens Netw*, vol. 15, no. 7, Jul. 2019, doi: 10.1177/1550147719865506.

[49] Y. Zhao, P. Tu, and M.-C. Chang, "Occupancy sensing and activity recognition with cameras and wireless sensors," in *Proceedings of the 2nd Workshop on Data Acquisition to Analysis*, 2019, pp. 1–6.

[50] S. Lu and E. C. Hameen, "Integrated IR Vision Sensor for Online Clothing Insulation Measurement," in *Proceedings of the 23rd Annual Conference of the Association for Computer-Aided Architectural Design Research in Asia*, 2018.

[51] J. Zou, Q. Zhao, W. Yang, and F. Wang, "Occupancy detection in the office by analyzing surveillance videos and its application to building energy conservation," *Energy Build*, vol. 152, pp. 385–398, Oct. 2017, doi: 10.1016/j.enbuild.2017.07.064.

[52] C. Gao, P. Li, Y. Zhang, J. Liu, and L. Wang, "People counting based on head detection combining Adaboost and CNN in crowded surveillance environment," *Neurocomputing*, vol. 208, pp. 108–116, 2016.

[53] O. K. T. Alsultan and M. T. Mohammad, "A Deep Learning-Based Assistive System for the Visually Impaired Using YOLO-V7," *Revue d'Intelligence Artificielle*, vol. 37, no. 4, pp. 901–906, Aug. 2023, doi: 10.18280/ria.370409.

[54] D. Shen, C. Ning, Y. Wang, W. Duan, and P. Duan, "Smart lighting control system based on fusion of monocular depth estimation and multi-object detection," *Energy Build*, vol. 277, Dec. 2022, doi: 10.1016/j.enbuild.2022.112485.

[55] "Performance Metrics Deep Dive - Ultralytics YOLO Docs." Accessed: Jun. 29, 2025. [Online]. Available: https://docs.ultralytics.com/guides/yolo-performance-metrics/

[56] J. Caballero-Peña, G. Osma-Pinto, J. M. Rey, S. Nagarsheth, N. Henao, and K. Agbossou, "Analysis of the building occupancy estimation and prediction process: A systematic review," Jun. 15, 2024, *Elsevier Ltd*. doi: 10.1016/j.enbuild.2024.114230.

[57] R. C. Navarro *et al.*, "Indoor occupancy estimation for smart utilities: A novel approach based on depth sensors," *Build Environ*, vol. 222, Aug. 2022, doi: 10.1016/j.buildenv.2022.109406.

[58] K. Sun, "'A Succinct Summary of Vision-based Building Occupancy Estimation,'" *Advancements in Civil Engineering & Technology*, vol. 5, no. 3, Feb. 2023, doi: 10.31031/acet.2023.05.000615.

[59] A. N. Sayed, Y. Himeur, and F. Bensaali, "Deep and transfer learning for building occupancy detection: A review and comparative analysis," Oct. 01, 2022, *Elsevier Ltd*. doi: 10.1016/j.engappai.2022.105254.