



University of
Nottingham
UK | CHINA | MALAYSIA

Application of Vision-Based Deep Learning Method for Building Occupancy and Thermal Comfort Predictions

Wuxia Zhang

Thesis submitted to the University of Nottingham for

The degree of Doctor of Philosophy

Jan 2025

ABSTRACT

Buildings are responsible for approximately 40% of global energy consumption and 30% of greenhouse gas emissions, highlighting their central role in sustainability efforts. Heating, ventilation, and air-conditioning (HVAC) systems are typically operated based on fixed schedules or occupant input, often misaligned with actual occupancy patterns. This mismatch leads to energy inefficiencies and occupant discomfort. Vision-based sensing technologies, particularly those employing cameras, offer the potential for real-time occupancy detection, providing a foundation for more adaptive building control strategies.

This thesis investigates the integration of vision-based deep learning methods for occupancy prediction and personalised thermal comfort modelling, aiming to enhance building energy performance and occupant well-being. Following a comprehensive review of the literature and identification of key research gaps, three main studies were conducted.

The first study develops and compares eight deep learning algorithms for vision-based occupancy detection. Annotated datasets were created from images collected in controlled environments. The models were evaluated using detection accuracy, mean average precision (mAP), and inference speed. Among the models tested, YOLOv8x achieved the highest accuracy (F1 score 0.87), while YOLOv8n offered a balance between accuracy and processing speed. When integrated into building simulations using IESVE, the predicted occupancy profiles led to significant improvements in HVAC energy estimation. A daily heating demand deviation of 13.45% in the base case was reduced to under 7% using deep learning models.

The second study compares the performance of thermal and standard RGB cameras for occupancy detection. Both modalities achieved comparable accuracy—around 70% with YOLOv8 and 80% with YOLOv10—given adequate training data. RGB cameras provide high-resolution detail but are susceptible to privacy concerns and visual interference. Thermal cameras, while offering better privacy and low-light performance, face limitations in scenarios involving overlapping occupants and residual heat. The results support thermal imaging as a viable, privacy-preserving alternative in suitable contexts.

The third study proposes a vision-based thermal comfort prediction model using deep learning and thermal imagery, offering an alternative to the conventional Predicted Mean Vote (PMV) approach. Personalised models achieved up to 68.49% accuracy in intra-subject tests, indicating potential for individual comfort prediction. However, reduced performance in cross-subject testing underscored the challenge of generalising thermal comfort models across diverse users.

Through systematic evaluation of algorithms, camera types, and comfort prediction strategies, this research advances the development of intelligent building systems. The findings suggest that vision-based approaches can support real-time, occupant-centred control of HVAC systems, contributing to improved energy efficiency and thermal comfort. Future work should focus on expanding datasets, refining model generalisability, and validating performance in real-world conditions.

ACKNOWLEDGEMENTS

The completion of this thesis marks the culmination of an extraordinary journey, one that would not have been possible without the guidance, encouragement, and support of many individuals for whom I am deeply grateful.

First and foremost, I extend my profound gratitude to my principal supervisor, Dr. John Kaiser Calautit, whose unwavering support, expert guidance, and insightful feedback have been the cornerstone of my academic journey. His dedication to my success and his belief in my potential provided me with the motivation to persevere through the most challenging moments of my research. I am equally grateful to my second supervisor, Prof. Yupeng Wu, for his invaluable advice and constructive critiques, which have greatly enriched this work.

A heartfelt thank you goes to my friends and college peers, whose invaluable assistance with my experiments made this work possible. Your willingness to support my research with your time and efforts, combined with the encouragement and camaraderie you provided, has left a lasting impact on my journey. I am deeply grateful for your contributions, which extended far beyond the academic realm.

Above all, I am profoundly grateful to my parents. Their unconditional love, encouragement, and sacrifices have been the foundation of my education and achievements. Without their support and belief in me, this journey would not have been possible. This thesis is as much their achievement as it is mine.

PREFACE

The work presented in this thesis was carried out between October 2020 and July 2024 at the Department of Architecture and Built Environment, University of Nottingham. As an outcome of the work, the following research outputs have been published:

Peer-reviewed journals:

Zhang W, Calautit J, Tien PW, Wu Y, Wei S. Deep learning models for vision-based occupancy detection in high occupancy buildings. *Journal of Building Engineering*. 2024; 98:111355

Wei S, Tien P, **Zhang W**, Wei Z, Wang Z, Calautit J. Deep Vision based detection for energy-efficiency and indoor air quality enhancement in highly polluted spaces. *Journal of Building Engineering*. 2024:108530.

Wang Z, Calautit J, Tien PW, Wei S, **Zhang W**, Wu Y, et al. An occupant-centric control strategy for indoor thermal comfort, air quality and energy management. *Energy and Buildings*. 2023; 285:112899

W. Zhang, Y. Wu, and J. K. Calautit, A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment, *Renewable and Sustainable Energy Reviews*, vol. 167, p.112704, 2022, <https://doi.org/10.1016/j.rser.2022.112704>.

Yan F, Shen J, **Zhang W**, Ye L, Lin X. A review of the application of green walls in the acoustic field. *Building Acoustics*. 2022;29(2):295-313. doi:10.1177/1351010X221096789

W. Zhang and J. Calautit, Occupancy behaviour and patterns: Impact on energy consumption of high-rise households in southeast China, *Smart Energy*, vol. 6, p. 100072, 2022, <https://doi.org/10.1016/j.segy.2022.100072>

Peer-reviewed conference papers:

W. Zhang, J.K. Calautit, Y. Wu, A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment, *16th Sustainable Development of Energy, Water and Environment Systems SDEWES Conference*, Dubrovnik, 10th-15th, Oct 2021.

W. Zhang, J.K. Calautit, N. Hamaza, Occupant Lifestyle: Impact on Energy Use of High-Rise Dwellings in Southeast China, *16th Sustainable Development of Energy, Water and Environment Systems SDEWES Conference*, Dubrovnik, 10th-15th, Oct 2021.

W. Zhang, J.K. Calautit, Y. Wu, Building Occupancy Prediction Through Machine Learning for Enhancing Energy Efficiency, Air Quality and Thermal Comfort: Review and Case Study, *International Conference on Energy and AI*, Aug 10-12, 2021, London, UK.

W. Zhang, P.W. Tien, J.K. Calautit, Y. Wu, Building Occupancy Prediction Through Machine Learning for Enhancing Energy Efficiency, Air Quality and Thermal Comfort: Review and Case Study, *Applied Energy Symposium 2021: Low carbon cities and urban energy systems*, August 24-27, 2021, Tokyo, Japan.

W. Zhang, J.K. Calautit, N. Hamaza, The Impact of Occupancy Energy Use Behaviour of High-Rise Dwellings in Southeast China, *Applied Energy Symposium 2021: Low carbon cities and urban energy systems*, August 24-27, 2021, Tokyo, Japan.

W. Zhang, P.W. Tien, J.K. Calautit, Y. Wu, Building Occupancy Prediction Through Machine Learning for Enhancing Energy Efficiency, Air Quality and Thermal Comfort: Review and Case Study, *International Conference on Sustainable Technology and Development*, 31 Oct-3 Nov 2021, Shenzhen, China.

W. Zhang, PW. Tien, J.K. Calautit, Y. Wu, F. Yan, Building Occupancy Prediction Through Machine Learning for Enhancing Energy Efficiency, Air Quality and Thermal Comfort: Review and Case Study, *Applied Energy Symposium 2022: Clean Energy towards Carbon Neutrality (CEN2022)*, April 23-25, 2022, Ningbo, China.

W. Zhang, PW. Tien, J.K. Calautit, Y. Wu, State-of-the-art deep learning and vision-based occupancy detection models for highly variable and high occupancy buildings, *International Conference of Net Zero Carbon Built Environment*, Nottingham, UK, July 3-5, 2024.

W. Zhang, J.K. Calautit, Y. Wu, A novel dynamic thermal comfort model in buildings with a vision-based deep learning method by using thermographic imaging, *SET2024: The 21st International Conference on Sustainable Energy Technologies*, Shanghai, China, August 12-14, 2024.

The details of the contribution of these works in this study are in 1.1.1Appendix.E.

CONTENTS

| | |
|--|-------|
| Abstract..... | i |
| Acknowledgements | iii |
| Preface | iv |
| Contents..... | vii |
| List of Figures..... | x |
| List of Tables | xvi |
| Nomenclature | xviii |
| Abbreviations | xx |
| 1. Introduction..... | 1 |
| 1.1 Research Background | 1 |
| 1.2 Aim and Objectives | 8 |
| 1.3 Thesis Outline | 9 |
| 2. Literature Review..... | 10 |
| 2.1 Method and commonly used occupancy prediction workflow based on ML | 12 |
| 2.1.1 The application of reviewed research..... | 13 |
| 2.1.2 The regions of reviewed studies | 14 |
| 2.2 Data collection for occupancy data..... | 16 |
| 2.2.1 Data collection, methods and privacy preservation..... | 16 |
| 2.2.2 Direct occupancy counting sensing technology | 20 |
| 2.2.3 Environmental sensors for data collection | 23 |
| 2.2.4 Data mining technologies | 27 |
| 2.3 Machine learning algorithms and their applications..... | 27 |
| 2.3.1 The trends of machine learning and deep learning..... | 29 |
| 2.3.2 Occupancy prediction..... | 30 |
| 2.3.3 Indoor air quality (IAQ) prediction | 38 |

| | | |
|-------|--|-----|
| 2.3.4 | Thermal comfort prediction..... | 40 |
| 2.3.5 | Energy consumption prediction..... | 42 |
| 2.4 | Validation of the prediction models: case study and time series..... | 46 |
| 2.5 | Thermal Comfort: Theories, Assessment, and Prediction | 50 |
| 2.6 | Research gap | 58 |
| 2.7 | Summary | 59 |
| 3. | Methodology | 62 |
| 3.1 | Research Structure | 62 |
| 3.2 | Experimental Setting..... | 64 |
| 3.3 | Ethical Considerations | 66 |
| 4. | Vision-based deep learning model comparison for occupancy detection..... | 66 |
| 4.1 | Introduction..... | 67 |
| 4.2 | Vision-based occupancy prediction method | 68 |
| 4.2.1 | Dataset generation | 68 |
| 4.2.2 | Deep learning model training and testing..... | 71 |
| 4.2.3 | Case study lecture room, testing and BES modelling | 74 |
| 4.3 | Experiment results and discussion..... | 78 |
| 4.3.1 | Comparison of state-of-the-art deep learning models in occupancy detection | 79 |
| 4.3.2 | Evaluation of the model performance in the case study building | 85 |
| 4.3.3 | Energy and CO ₂ simulation results..... | 94 |
| 4.4 | Summary | 99 |
| 5. | Occupancy prediction performance comparison of standard camera and thermographic imaging..... | 102 |
| 5.1 | Introduction..... | 102 |
| 5.2 | Occupancy prediction methodology | 106 |
| 5.2.1 | Case study experiment setups..... | 107 |
| 5.2.2 | Training dataset generation | 110 |
| 5.2.3 | Training model | 113 |
| 5.3 | Experiment Results | 117 |
| 5.3.1 | Video inference results | 117 |
| 5.3.2 | Different scenarios result in Cross-Video Experiments | 123 |

| | | |
|-------|---|-----|
| 5.4 | Discussion..... | 133 |
| 5.4.1 | Metrics evaluation | 133 |
| 5.4.2 | Detection performance comparison in simple and complex scenarios | 137 |
| 5.4.3 | Additional findings in vision-based occupancy prediction | 138 |
| 5.5 | Summary | 143 |
| 6. | Dynamic thermal comfort prediction with thermographic imaging | 146 |
| 6.1 | Introduction..... | 146 |
| 6.2 | Thermal comfort prediction method | 148 |
| 6.2.1 | Case study room setup..... | 151 |
| 6.2.2 | Experimental setup and procedure | 153 |
| 6.2.3 | Thermal camera calibration experiment | 157 |
| 6.2.4 | Deep learning model..... | 160 |
| 6.3 | Results and discussions..... | 168 |
| 6.3.1 | Comparison between PMV and TSV | 168 |
| 6.3.2 | Deep learning prediction of thermal comfort based on the intra-subject dataset | 173 |
| 6.3.3 | Deep learning-based detection of thermal comfort based on cross-subject dataset | 180 |
| 6.4 | Summary | 185 |
| 7. | Conclusion and Future Work | 188 |
| 7.1 | Conclusions..... | 188 |
| 7.2 | Contribution to Knowledge | 192 |
| 7.3 | Overall Study Limitations..... | 195 |
| 7.4 | Recommendations for Future Work | 197 |
| | References | 200 |
| | Appendices | 220 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1-1 The difference between the black box model, white box model and grey box model..... | 3 |
| Figure 2-1 An overview of the application of machine learning in the built environment based on the reviewed studies from 2011 to 2021..... | 13 |
| Figure 2-2 The location of case studies in the reviewed papers conducted from 2011 to 2021..... | 14 |
| Figure 2-3 The typical procedure of occupancy prediction with machine learning, validation and applications in the built environment. | 16 |
| Figure 2-4 The proportion of building types in the reviewed case studies | 17 |
| Figure 2-5 Case study building types in different regions of reviewed studies..... | 18 |
| Figure 2-6 Data collection methods and their related application in the reviewed studies. | 20 |
| Figure 2-7 Summary of the reviewed studies from 2011 to 2021 using machine learning algorithms..... | 30 |
| Figure 2-8 The existing workflow of IAQ and thermal comfort prediction and the potential improvement..... | 42 |
| Figure 2-9 The implementation scale of different applications of prediction models in reviewed studies..... | 48 |
| Figure 2-10 The prediction timeframe and experimental method were conducted in different regions based on the reviewed studies..... | 49 |

| | |
|---|----|
| Figure 2-11 Comparison between workflow steps of machine learning and deep learning for vision-based thermal comfort prediction models. | 57 |
| Figure 3-1 The workflow for PhD Methodology in this thesis. | 63 |
| Figure 4-1 The workflow of the proposed vision-based deep learning method for occupancy detection. | 68 |
| Figure 4-2 Example images from the training dataset, showcasing humans in various environments including classrooms, offices, and outdoor scenes. The diversity of the dataset ensures the generalisation of the deep learning models for occupancy detection in different settings. | 71 |
| Figure 4-3 (a) The Marmont Centre at the University of Nottingham, UK. (b) The indoor view of the case study lecture room. (c) The floor plan and installed sensors layout of the case study building (d) The picture of the camera in this test. (e) The Awair Element environmental sensors in indoor and outdoor | 75 |
| Figure 4-4 The workflow for the different cameras employed in the case study experiment on Dec 2 nd , 2022. | 77 |
| Figure 4-5 The frame at 15:15:10 compares the deep learning models' detection of the participants entering the room. | 80 |
| Figure 4-6 The frame at 15:45:00 compares the deep learning model detection of the participants in the middle of the lecture. | 82 |
| Figure 4-7 The frame at 16:09:40 compares the deep learning models' detection of participants leaving the room. | 84 |
| Figure 4-8 Examples of TP, FP and FN in a frame taken from the result of Faster R-CNN, YOLOv5n, YOLOv7 and YOLOv8n at 15:35:00. | 86 |

| | |
|--|-----|
| Figure 4-9 Performance of deep learning models comparing training time, inference time and accuracy (higher accuracy model occupied bigger circular area)..... | 90 |
| Figure 4-10 The occupancy profiles predicted by the deep learning models, compared to the ground truth. | 91 |
| Figure 4-11 A frame taken from the YOLOv8n inference detection videos comparing 4 different cameras at 15:45:00..... | 92 |
| Figure 4-12 Occupancy profiles predicted by the deep learning models based on the different detection camera locations using YOLOv8n compared to ground truth. | 94 |
| Figure 4-13 Predicted vs. recorded data for a) CO ₂ concentration, b) internal gains, and c) heating loads. | 96 |
| Figure 4-14 The predicted heating energy of the case study room on Dec 2 nd , 2022, based on the simulation of deep learning model profiles, the “Ground Truth” profile and the “Base case” profile. | 98 |
| Figure 5-1 Evolution of the use of thermal imaging across occupancy studies, from low to medium pixel grids (Kraft et al., 2021, Sirmacek and Riveiro, 2020) to higher detail (present). | 105 |
| Figure 5-2 The workflow for the comparison of standard and thermal cameras in vision-based occupancy detection..... | 106 |
| Figure 5-3 The details of the buildings where the field studies are conducted..... | 108 |
| Figure 5-4 Example images from datasets 1, 2 and 3. Dataset 1 represents simple scenarios and 2 and 3 for crowded ones. | 112 |

| | |
|---|-----|
| Figure 5-5 The normalized confusion matrix for the Same-Video Experiment, Split-Video Experiment and Cross-Video Experiments for (a) standard and (b) thermal datasets..... | 116 |
| Figure 5-6 The generated occupancy profiles in each experiment. (a) Same video, (b) Split video, and Cross-Video (c) Experiment 1, (d) Experiment 2, (e) Experiment 3, (f) Experiment 4. Three dark grey bars indicate the boundary of four stages of the inference video, which are entering the room, discussion (sitting), leaving the room, and discussion (standing). | 121 |
| Figure 5-7 The examples of inference video. From Stages 1 to 4: entering the room, discussion (sitting), leaving room, and discussion (standing). | 123 |
| Figure 5-8 The Cross-Video Experiments video inference results in comparison for the standard and thermal models in the entering room stage. | 124 |
| Figure 5-9 The detection results of thermal models at 01:14 of the inference video in Experiment 1 to 4. | 126 |
| Figure 5-10 The Cross-Video experiments video inference results in comparison for standard and thermal models in the discussion (sitting) stage. | 128 |
| Figure 5-11 The Cross-Video experiments video inference results in comparison for standard and thermal models in the leaving room stage. | 130 |
| Figure 5-12 The Cross-Video experiments video inference results in comparison for standard and thermal models in the discussion (standing) stage. | 132 |
| Figure 5-13(a) The comparison for mAP and manual counted accuracy in all experiments. (b) The accuracy comparison of simple and complex scenarios in different experiments. | 135 |

| | |
|---|-----|
| Figure 5-14 (a) Examples of occupancy detection in previous studies, and (b) example pictures of deep learning model detecting occupants in the present study. The thermal imprints left on chairs: (c) the deep learning model mis-detected it as an occupant in Cross-Video Experiment 3, (d) additional images were added to the dataset with thermal imprints left on chairs, and (e) the model correctly ignored the thermal imprints left on chairs. | 140 |
| Figure 5-15 The detection of overlapping people. The dataset of Experiment 2 added more images with people in different environment and Experiment 3 added crowded scenarios specifically. | 141 |
| Figure 6-1 The workflow for the intra-subject thermal comfort detection model evaluation. | 150 |
| Figure 6-2 Detailed workflow for the proposed vision-based thermal comfort detection method. | 151 |
| Figure 6-3 (a) The Sustainable Research Building and (b) overview of the case study room. | 153 |
| Figure 6-4 The thermal camera and environment sensors' location in the case study room. | 155 |
| Figure 6-5 Experimental setup for thermal camera calibration using high-temperature type k thermocouples. | 158 |
| Figure 6-6 Comparing thermal camera and thermocouple measurements over time. | 159 |
| Figure 6-7 The architecture of the YOLOv8 algorithm, which is divided into four parts, including backbone, neck, head, and loss (Ju and Cai, 2023). | 162 |

| | |
|--|-----|
| Figure 6-8 Example pictures in the dataset of different classification for subject 3 and subject 7..... | 163 |
| Figure 6-9 The PMV values over time in 14 tests. | 171 |
| Figure 6-10 Comparison of a) TSV against temperature and b) TSV against PMV in 14 experiments. | 172 |
| Figure 6-11 Comparison between TSV, PMV and deep-learning model results for 14 subjects. | 174 |
| Figure 6-12 The screenshots of the validation video from 15:30-15:35 for every 1 minute in the experiment of Subject 2. The first number indicates the TSV, and the second number is the PMV. | 176 |
| Figure 6-13 The result comparison of the cross-subject models with corresponding TSV and PMV. | 181 |
| Figure 6-14 The screenshots of results in a cross-subject test for Subject 5 with the dataset from Subjects 1 to 4 and 7 to 14. | 184 |

LIST OF TABLES

| | |
|--|----|
| Table 2-1 Information on existing reviews in recent years. | 11 |
| Table 2-2 Comparison and key findings between different direct occupancy counting methods in recent studies. | 22 |
| Table 2-3 Recent studies on occupancy detection using environmental sensors. | 24 |
| Table 2-4 The information about studies using various algorithms in occupancy state/number/activities prediction. | 32 |
| Table 2-5 Examples of research work on occupancy detection and prediction using the vision-based method | 34 |
| Table 2-6. Summary of the commonly used machine learning algorithms for different applications | 44 |
| Table 2-7. Summary of the algorithm, prediction time and accuracy in some of the reviewed studies. | 50 |
| Table 2-8 Vision-based machine learning research for building thermal comfort in recent years. | 54 |
| Table 3-1 Summary of Experimental Settings and Methodology..... | 65 |
| Table 4-1 Dataset creation and preprocessing steps for occupancy detection | 70 |
| Table 4-2 Comparison of different objection detection models' training performance in this study. The best results for each category are highlighted in bold. | 73 |

| | |
|--|-----|
| Table 4-3 Information of the case study lecture room and occupancy profiles. | 75 |
| Table 4-4 Environmental sensors and cameras used in the case study experiment. | 76 |
| Table 4-5 IES Modelling construction details including U-values (W/m ² K) and thickness. | 78 |
| Table 4-6 Comparison of the performance of the deep learning models during three distinct phases: as participants enter the room, during the lecture, and as they exit the room..... | 89 |
| Table 4-7 The model performance across all four cameras. | 94 |
| Table 5-1 Details of the setting of experiments conducted for this study. | 109 |
| Table 5-2 The details of the different datasets used as Cross-video experiments in this study. | 111 |
| Table 5-3 The training details of all deep learning tests in this chapter. | 114 |
| Table 5-4 The detailed accuracy results for all experiments with video inference. | 118 |
| Table 6-1 Sensors for collecting environmental data in the field experiment..... | 155 |
| Table 6-2 Detailed information about the deep learning model for each subject..... | 164 |
| Table 6-3 The training details for the multi-people dataset. | 167 |
| Table 6-4 The MAE and RMSE of the PMV and deep learning method results for 14 subjects. | 178 |

NOMENCLATURE

| | | |
|-----------|--------------------------------------|-----------------------|
| h_c | Convective Heat Transfer Coefficient | W/(m ² ·K) |
| I_{cl} | Clothing Insulation | clo |
| P_a | Vapour Pressure | kPa |
| T_{cl} | Temperature of The Clothing Surface | °C |
| T_{mr} | Mean Radiant Temperature | °C |
| f_{cl} | Clothing Factor | W/(m ² ·K) |
| D | Diameter | mm |
| dpi | Dots per Inch | dots/inch |
| M | Metabolic Rate | W/m ² |
| mAP | Mean Average Precision | % |
| MIT | Mean Inference Time | s |
| $PM\ 2.5$ | Particulate Matter 2.5 | µg/m ³ |
| RH | Relative Humidity | % |
| $RMSE$ | Root Mean Square Error | - |
| $RMSPE$ | Root Mean Squared Percentage Error | % |
| SET | Standard Effective Temperature | °C |

| | | |
|------------|----------------------------|---------------------------|
| T_a | Air Temperature | °C |
| T_g | Globe Temperature | °C |
| V | Air Velocity | m/s |
| $VOCs$ | Volatile Organic Compounds | ppb |
| W | Work | W/m ² |
| ϵ | Emissivity | ranging from 0 to 1 |

ABBREVIATIONS

| | |
|---------------|--|
| <i>AdB</i> | AdaBoost |
| <i>AI</i> | Artificial Intelligence |
| <i>ANFIS</i> | Adaptive Neuro-Fuzzy Interference System |
| <i>ANN</i> | Artificial Neural Network |
| <i>ASHRAE</i> | American Society of Heating Refrigerating and Airconditioning Engineer |
| <i>BES</i> | Building Energy Simulation |
| <i>CIBSE</i> | Chartered Institution of Building Services Engineers |
| <i>CNN</i> | Convolutional Neural Network |
| <i>CV</i> | Computer Vision |
| <i>DBF</i> | Deepsort, dynamic Bayesian fusion |
| <i>DCM</i> | Data Collection Method |
| <i>DNNs</i> | Deep Neural Networks |
| <i>DT</i> | Decision Tree |
| <i>FFNN</i> | Feed Forward Neural Network |
| <i>FN</i> | False Negative |
| <i>FP</i> | False Positive |
| <i>GAN</i> | Generative Adversarial Network |

| | |
|---------------|--|
| <i>GB</i> | Gradient Boosting |
| <i>HMI</i> | Human Machine Interface |
| <i>HMM</i> | Hidden Markov Model |
| <i>HOG</i> | Histogram of Oriented Gradients |
| <i>HVAC</i> | Heating, Ventilation and Air-Conditioning |
| <i>IAQ</i> | Indoor Air Quality |
| <i>ICT</i> | Information and Communication Technology |
| <i>IES VE</i> | Integrated Environmental Solutions Virtual Environment |
| <i>IoT</i> | Internet of Things |
| <i>IoU</i> | Intersection over Union |
| <i>KNN</i> | K-Nearest Neighbour |
| <i>LMSR</i> | Linear Model Stepwise Regression |
| <i>LR</i> | Logistic Regression |
| <i>LSTM</i> | Long Short-Term Memory |
| <i>ML</i> | Machine Learning |
| <i>MLP</i> | Multilayer Perceptron |
| <i>MPC</i> | Model Predictive Control |
| <i>MSE</i> | Mean Square Error |
| <i>MPMV</i> | Metabolic Predicted Mean Vote |

| | |
|--------------|---|
| <i>NNARX</i> | Nonlinear Autoregressive Network with Exogenous |
| <i>NB</i> | Naïve Bayes |
| <i>PV</i> | Solar Photovoltaic |
| <i>PMV</i> | Predicted Mean Vote |
| <i>PIR</i> | Pyroelectric Infrared |
| <i>RBFN</i> | Radial Basis Function Network |
| <i>R-CNN</i> | Regions with Convolutional Neural Networks |
| <i>RF</i> | Random Forest |
| <i>RFID</i> | Radio Frequency Identification Devices |
| <i>RoI</i> | Regions of Interest |
| <i>RPN</i> | Region Proposal Network |
| <i>SOM</i> | Self Organizing Map |
| <i>SSD</i> | Single Shot Detector |
| <i>SVC</i> | Support Vector Classification |
| <i>SVM</i> | Support Vector Machine |
| <i>SVR</i> | Support Vector Regression |
| <i>TN</i> | True Negative |
| <i>TP</i> | True Positive |
| <i>YOLO</i> | You Only Look Once |

1. INTRODUCTION

1.1 Research Background

Buildings are responsible for up to 40% of the global total energy (Cao et al., 2016) and 30% of greenhouse gas (Sbeci, 2009). Buildings consume a lot of energy since they serve various purposes and consume energy (Bosák and Palko, 2014). Particularly, buildings now combine traditional energy services systems like heating, ventilation, and air conditioning (HVAC), lighting, power distribution, and water systems with on-site power-generating systems like solar photovoltaic (PV), wind turbines, and electric vehicle charging systems (Šimić and Devedžić, 2003). At the same time, people spend 80%–90% of their time indoors, and thermal comfort is a critical factor for physical health, mental well-being, and productivity (Mujan et al., 2019). The main challenge is to find a balance between providing a comfortable and healthy indoor environment and minimising the energy demand.

Despite the massive quantity of energy used by buildings, thermal comfort is not always achieved. The study showed that in a conditioned office building, 75% of occupants report that they are dissatisfied with their thermal comfort (Erickson and Cerpa, 2012). Another field study in the US indicated that only 60% of occupants in 60 office buildings were satisfied with their thermal environment (Karmann et al., 2017). Even high-performance and energy-efficient buildings may not be as comfortable or healthier than other buildings as they are intended to be (Roulet et al., 2006).

Energy consumption in buildings is influenced by many factors, including weather conditions, building design, HVAC system efficiency, and the operation of appliances. Among these, occupancy is one of the most complex and least predictable factors. Unlike other variables, occupancy is dynamic, varying in time, space, and behaviour (Yoshino et al., 2017). In the past, occupants' behaviours were observed (Barthelmes et al., 2016) or

through interviews and surveys (Rebaño-Edwards, 2007) to generate a fixed occupancy schedule (Tuohy et al., 2009) which can be used in building models or simulations for existing buildings. However, the actual occupancy behaviour is difficult to predict since it is time-varying and identity in different cases. Therefore, proposing a thorough and accurate occupancy prediction model is necessary for building energy conservation and to guide occupant behaviour modelling in building energy simulation (Nastasi et al., 2022).

As a result, static schedules can lead to inefficiencies, such as ventilation, heating, cooling, and lighting spaces that are unoccupied, or failing to adequately condition spaces that are occupied outside of expected times. This mismatch between predicted and actual occupancy not only wastes energy but also diminishes the comfort and satisfaction of building occupants (Pappalardo and Reverdy, 2020). Addressing these inefficiencies requires more dynamic and accurate methods of occupancy detection that can adapt to the real-time presence and movement of people within buildings.

In the last decade, new powerful tools, including machine learning methods and data mining techniques, have been suggested to diagnose unnoticed relationships and summarise the data in innovative ways according to large information datasets, as discussed in many studies (Nastasi et al., 2022). To better understand energy usage in buildings, research tends to study the diversification of occupancy schedules based on big data streams (Ding et al., 2021a). A lot of research has been conducted to bridge the gap between occupancy prediction and building control while maintaining thermal comfort, which naturally has a significant impact on building energy use. One research with an AI-based method achieved energy conservation of up to 30% by using occupancy and eight different physical sensors (Turley et al., 2020). Another paper proposed an integrated framework for an HVAC system that suggested a significant reduction in comfort dissatisfaction, going from 25% with the baseline strategy to 0% dissatisfaction while decreasing the energy cost by more than 10% (Winkler et al., 2020).

Many occupancy models have been created over the last twenty years to simulate occupant unpredictability and variety and generate stochastic occupancy models for making accurate simulations (Kamel et al., 2020). The three types of prediction models are the physical model or white-box model, the black-box model, also called the data-driven model, and the grey model (Foucquier et al., 2013) as shown in Figure 1-1. White-box models produce detailed simulations of a building's energy performance, with details such as the building material, HVAC control, and management systems (Coakley et al., 2014). In addition, creating a white-box model takes time and some building details are difficult to obtain. Data-driven models are fast to construct and provide acceptable results with good data quality, but they require a large amount of data, and their parameters and inputs have no obvious physical meaning (Meng et al., 2020). Mixture models combine physical and data-driven models, inheriting the advantages and disadvantages of both techniques. Traditional energy models with sets of specified static coefficients multiplied by a maximum room occupancy were white-box models with extensive building information and certain occupancy characteristics (Abushakra et al., 2004).

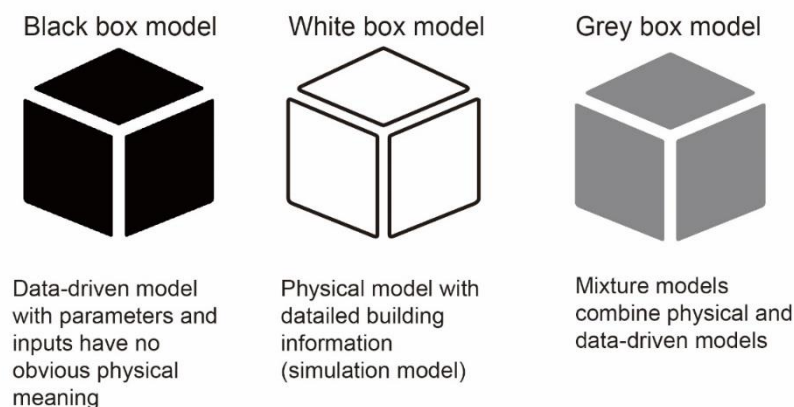


Figure 1-1 The difference between the black box model, white box model and grey box model.

With the rapid advancement of computer technology, data-driven approach (black-box) models have shown great potential in building energy models to simulate and predict

related appliances, including occupancy behaviour, thermal comfort, IAQ and energy consumption. A study compared the occupancy prediction model with and without a machine learning algorithm and showed that the accuracy was significantly improved and 30% energy saving can be achieved with the proposed algorithm (Aftab et al., 2017). Another study using a learning-based model predictive control (MPC) technique achieved significant energy savings, with 40.56% less cooling and 16.73% less heating power while keeping occupants comfortable (Eini and Abdelwahed, 2019).

Traditionally, occupancy detection methods have relied on a variety of sensors such as Passive Infrared (PIR) sensors (Sheikh Khan et al., 2021), carbon dioxide (CO₂) sensors (Franco and Leccese, 2020), Radio-Frequency Identification (RFID) systems (Li et al., 2012), and electricity meters (Razavi et al., 2019). These technologies can provide useful data but often fall short in accuracy and granularity, especially in high occupancy environments where the number of occupants and their movements can vary significantly. Furthermore, these sensor-based methods can be intrusive and may not effectively capture the dynamic nature of human occupancy.

More recent approaches have incorporated Wi-Fi signals (Alishahi et al., 2022) and cameras (Tien et al., 2022) to monitor occupancy. While Wi-Fi-based methods can offer improved coverage, they still struggle with accuracy and real-time responsiveness. Cameras, on the other hand, present a promising solution by using visual data to track occupancy more precisely (Gao et al., 2022). Cameras continuously monitor visual cues to track people's presence, movements, and interactions, which is crucial in dynamic, crowded settings. This is particularly valuable in environments where understanding occupancy patterns and behaviours is essential.

Earlier computer vision methods (Tong et al., 2013) for occupancy detection often required extensive computational resources and were hindered by the limitations of available hardware and algorithms. These methods typically involved complex feature

extraction and pattern recognition, which were not robust enough for real-time applications and, hence, were not scalable or efficient for widespread adoption.

In recent years, improvements in computational power, alongside the advancements in computer vision and deep learning have opened new avenues for occupancy detection. Vision-based approaches offer a non-intrusive and highly accurate means of monitoring occupancy by analysing visual data captured from cameras. Deep learning, particularly Convolutional Neural Networks (CNNs), has transformed the field of computer vision (O'Mahony et al., 2020). CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images. This ability to learn and extract intricate patterns and features makes CNNs particularly effective for tasks such as object detection and recognition, which are critical for accurate occupancy detection (Sindagi and Patel, 2018).

Studies have shown that the CNN method performs particularly well in high-density scenes, making it suitable for environments where the number of occupants can vary widely and rapidly. This enhanced capability is due to CNNs' proficiency in handling occlusions and complex visual data, allowing for more reliable detection and tracking of individuals in crowded settings. Specifically, models such as Single Shot MultiBox Detector (SSD), Faster Region-based Convolutional Neural Networks (Faster R-CNN), and You Only Look Once (YOLO) have shown great promise.

These models enhance occupancy detection by processing large amounts of visual data to identify and count occupants in real-time. SSD and YOLO are known for their speed and efficiency, making them suitable for applications requiring real-time analysis. Faster R-CNN, while slightly slower, provides higher accuracy and is effective in complex scenes with varying levels of occupancy. Despite their promising performance in controlled environments, the effectiveness of these methods in real-world building environments has not been thoroughly examined, particularly in situations where the

presence of people is essential for building energy management (Li et al., 2018b). Further research is needed to validate the performance of deep learning models in real-world building scenarios and to understand their impact on energy efficiency.

Within vision-based systems, cameras play a central role, with two primary technologies being thermal cameras and standard cameras (Kim et al., 2023). Thermal cameras detect heat signatures, making them effective in low-light environments and providing privacy advantages by avoiding identifiable facial features (Gade and Moeslund, 2014). However, they are often more expensive and have lower spatial resolution compared to standard cameras. Standard cameras, on the other hand, capture visible light, offering higher resolution and widespread availability at lower costs (Lydon et al., 2019). They can support detailed analyses of occupancy but are less effective in low-light conditions and raise privacy concerns (Agrawal et al., 2022). Comparative studies of thermal and standard cameras, especially in occupancy prediction, remain limited, leaving a critical gap in the understanding of their effectiveness.

Traditional HVAC systems, which operate at fixed settings irrespective of occupancy or environmental changes, often result in energy inefficiencies, such as over-conditioning unoccupied spaces. In contrast, advanced HVAC systems equipped with real-time monitoring capabilities can dynamically adapt to changing conditions by analysing factors such as occupancy, temperature, and humidity (Kim and Hong, 2020). This responsiveness not only reduces unnecessary heating or cooling in underutilised spaces but also ensures optimal comfort for occupants. As a result, such systems provide a balanced approach to energy management, enhancing efficiency without compromising user satisfaction (Lan et al., 2010). Studies indicate that these adaptive systems can achieve up to 21.4% energy savings compared to static HVAC setups, highlighting their potential for mitigating energy waste while improving indoor environmental quality (Jung and Jazizadeh, 2020).

To optimise both energy efficiency and occupant well-being, various thermal comfort models have been developed to regulate indoor climates. One of the most widely used models is the Predicted Mean Vote (PMV), which underpins many modern HVAC control strategies. PMV-based systems continuously monitor indoor environmental data, such as temperature and humidity, and adjust HVAC settings accordingly to create a more comfortable environment compared to traditional static HVAC systems (Mao et al., 2019). Beyond energy savings, PMV-based control enhances occupant comfort by dynamically fine-tuning HVAC operations to maintain a stable thermal environment (Choi et al., 2024).

The PMV model considers six key variables: air temperature, relative humidity, wind velocity, mean radiant temperature, metabolic rate, and clothing insulation (Fanger, 1970b). However, obtaining these variables can be both costly and challenging in real-world buildings. For instance, measuring mean radiant temperature and air velocity requires sophisticated and expensive instruments, which are often impractical for continuous monitoring across multiple building zones. Meanwhile, parameters such as clothing insulation and metabolic rates are typically assumed or simplified, as it is typically not practical to collect precise data in real-time for every occupant (d'Ambrosio Alfano et al., 2011). This reliance on estimated values can limit the effectiveness of PMV-based control, particularly in environments with diverse occupant profiles or rapidly changing conditions.

Moreover, the PMV model struggles to account for changes in thermal comfort during dynamic scenarios (Cheung et al., 2019b). It also overlooks individual differences, which can lead to inaccurate predictions of personal thermal comfort and inefficient energy use (Jazizadeh et al., 2014). These limitations highlight the need for more adaptive and personalised approaches to thermal comfort prediction that can respond to dynamic conditions and account for individual variability.

1.2 Aim and Objectives

The overall aim of this thesis is to explore the application of vision-based deep learning frameworks for improving occupancy prediction and thermal comfort modelling in building environments. Occupancy prediction plays a critical role in optimizing building energy management, as it directly affects the operation of HVAC, lighting, and other energy-intensive systems (Pang et al., 2023). However, traditional approaches often rely on static assumptions and generalised models, which fail to capture the complexity and variability of real-world occupancy (Esrafilian-Najafabadi and Haghighat, 2022). Similarly, thermal comfort is frequently assessed using generalised models, such as the PMV, that overlook individual differences and dynamic conditions (Dong et al., 2021). These limitations present an opportunity for advanced vision-based deep learning methods to bridge the gap between energy efficiency and occupant satisfaction.

By enabling real-time occupancy and personalised comfort prediction using non-intrusive vision-based sensing, the proposed framework can support the development of smart building systems that dynamically respond to occupants' presence and comfort needs. The integration of such models into building management systems (BMS) has the potential to improve energy efficiency, reduce HVAC-related emissions, and enhance indoor environmental quality. Moreover, the use of thermal cameras offers a privacy-conscious solution for sensitive environments, expanding the applicability of vision-based control in contexts such as offices, educational settings, and healthcare facilities.

To achieve the aim, the main objectives listed below are carried out in this research.

- 1) Conduct an in-depth and critical review of existing literature on machine learning applications in building systems, focusing on occupancy prediction, indoor air

quality (IAQ) prediction, thermal comfort modelling, and energy consumption optimization.

- 2) Collect, annotate, and test a dataset of occupant images from random and dynamic environments, ensuring diversity in environmental conditions, occupant demographics, and behaviours to support model development.
- 3) Evaluate the performance of various deep learning algorithms (e.g., SSD, Faster R-CNN, YOLO series) for real-time occupancy prediction, identifying the most effective method for optimizing energy use and occupant detection in complex environments.
- 4) Investigate and compare the performance of standard and thermal cameras for occupancy prediction, to systematically analyse trade-offs in privacy, cost, and spatial resolution for building energy systems.
- 5) Compare predicted occupancy data with ground truth (actual) measurements and evaluate the impact of accurate occupancy detection on energy use and CO₂ concentrations through building energy simulations.
- 6) Develop a personalized vision-based thermal comfort prediction model using real-time occupant data from thermal cameras combined with environmental conditions, addressing limitations of traditional comfort models and exploring applications for improving occupant well-being.

1.3 Thesis Outline

This thesis is organized into six chapters, the summary for each chapter is listed as follows:

Chapter 2 presents a comprehensive review of the literature, covering current approaches to occupancy prediction, especially the use of vision-based methods, and advancements in machine learning for thermal comfort assessment. Research gaps are identified to position the study within the existing body of knowledge.

Chapter 3 focuses on the comparison of eight deep-learning algorithms for occupancy prediction, evaluating their performance in detecting and counting occupants in a lecture room.

Chapter 4 presents a comparative study of thermal and standard cameras, analysing their respective strengths, limitations, and suitability for real-world applications in building energy management.

Chapter 5 explores thermal comfort prediction using thermal cameras and deep learning methods, emphasizing the development of personalized models that integrate real-time data for improved occupant satisfaction.

Chapter 6 concludes the thesis by summarizing the key findings, discussing the contributions to the field, and proposing future research directions.

2. LITERATURE REVIEW

Some work presented in this Chapter was previously published in the journal [Renewable and Sustainable Energy Reviews] as titled *A Review on Occupancy Prediction Through Machine Learning for Enhancing Energy Efficiency, Air Quality and Thermal Comfort in the Built Environment* by author Wuxia Zhang and co-authors Yupeng Wu and John Kaiser Calautit. I played a major role in Conceptualization, Methodology, and Writing - the original draft and this study were conceived by all the authors.

This chapter conducted an in-depth and critical evaluation of the application of machine learning on buildings including occupancy prediction, indoor air quality prediction, thermal comfort prediction and energy consumption prediction, specific to the vision-based method for occupancy prediction. In 2012, a brief review was conducted of the methods for predicting building energy consumption, including ANNs and SVM (Zhao and Magoulès, 2012). In 2021 a review compared the AI-based and conventional models employed in building energy consumption prediction with occupancy factors and proved that AI-based models had better accuracy (Ramokone et al., 2021). Another work reviewed studies on electrical load prediction and provided an overview of the prediction timescale and potential model solutions (Kuster et al., 2017). The use of machine learning in the various phases of the building lifecycle was examined, and research gaps in the design, construction, operation and maintenance, and control, were investigated in another paper (Hong et al., 2020). Most of these review papers focused on the occupancy detection approach and performance, while in terms of its application in buildings, most of the studies evaluated its impact on energy efficiency but not thermal comfort and air quality (as shown in Table 2-1). This work argues that the occupancy behaviour data obtained can be employed to minimise energy and at the same time provide a comfortable and healthy environment. For example, the occupancy prediction method can be integrated into a framework or model which can control and optimise the operation of the HVAC regarding energy, comfort and health.

Table 2-1 Information on existing reviews in recent years.

| Ref. | Year | Journal | Research Focus and Gaps |
|---------------------|-------------|-------------------------------|---|
| (Wei et al., 2019a) | 2019 | <i>Indoor Air</i> | Focused on the sensors collecting air quality index and have not considered the occupancy impact. |
| (Saha et al., 2019) | 2019 | <i>Energy & Buildings</i> | Focused on occupancy sensing review and lack of consideration of future prediction and validation methods |

| | | | |
|---|------|--|---|
| (Xilei et al., 2020) | 2020 | <i>Energy & Buildings</i> | Mainly focused on occupancy detection and estimation, not enough integrating occupancy information with models. |
| (Hong et al., 2020) | 2020 | <i>Energy & Buildings</i> | Examined papers using machine learning in different stages of the building life cycle. |
| (Yao and Shekhar, 2021) | 2021 | <i>Building and Environment</i> | Focused on the various types of MPC and their software implementation |
| (Ramokone et al., 2021) | 2021 | <i>Sustainable Energy Technologies and Assessments</i> | Focused on prediction of occupant number/level and fail to locate the impact of occupancy-interlinked inhabitant behaviour. |
| (Ding et al., 2021b) | 2021 | <i>Building Simulation</i> | Focuses on sensors and algorithms used in occupancy prediction and does not pay attention to the interaction of occupants with the building systems. |
| (Fu et al., 2021) | 2021 | <i>Renewable and Sustainable Energy Reviews</i> | Focused on the energy model but did not pay enough attention to the occupancy factors and their comfort. |
| (Esrafilian-Najafabadi and Haghighat, 2021) | 2021 | <i>Building and Environment</i> | Divided the occupancy prediction models into state/level prediction and occupancy activities prediction, but not much discussion about activities prediction. |

2.1 Method and commonly used occupancy prediction workflow based on ML

Although there is a large amount of literature on building occupancy prediction using machine learning and a great number of review articles, what is lacking is a straightforward categorization and organization of mathematic methodologies and technologies, allowing for the definition of a useful (or ideal) "occupancy data structure." Therefore, we consider articles published from 2011 to 2021 in the main databases such as Scopus and Thomas Reuters' Web of Science. The keywords included

“building, occupancy prediction, machine learning” & “thermal comfort, occupancy prediction, building”. The keywords “thermal comfort, machine learning, artificial intelligence, comfort factor, indoor air temperature, and control method” were also used to identify more related publications. We focus on papers that employ machine learning to predict occupancy in buildings and related applications. Review papers and irrelevant papers were excluded, for example, some research only focused on occupancy detection and was not suitable for the review purpose.

2.1.1 The application of reviewed research

160 papers were selected, and a timely review was proposed, which can help guide the future research of occupancy prediction with machine learning regarding building design, operation, and research activities and provide a better understanding of occupancy behaviour and building performance.

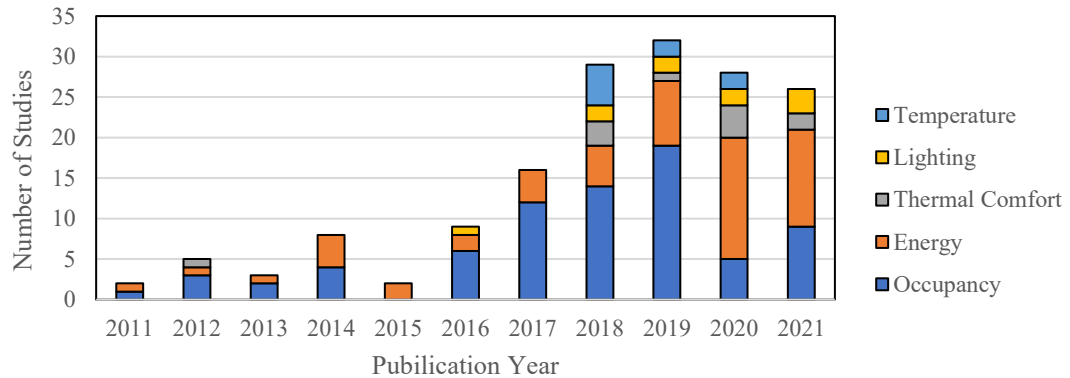


Figure 2-1 An overview of the application of machine learning in the built environment based on the reviewed studies from 2011 to 2021.

In general, the number of machine learning and its applications research in building environments is rising, particularly in recent years (Figure 2-1). These applications include the prediction of occupancy state, occupants' interactions with thermal comfort, energy consumption, indoor temperature, and lighting use. Occupancy state prediction

was the most popular application of machine learning models until 2020, while the number of studies on energy consumption prediction increased. This could be due to the development of prediction models, which can be specifically used for more detailed problems like the comfort state and the occupancy activities instead of just predicting if the room is occupied or not. Also, it shows an increasing awareness of energy efficiency and occupancy comfort in the built environment.

2.1.2 The regions of reviewed studies

The case studies in the reviewed papers were mostly conducted in three big geographic regions: Europe, North America, and Asia. Most of the early studies were in Europe and North America, while studies in Asia have increased since 2016, as shown in

Figure 2-2. In recent years when the topic became more popular, these three main regions dominated this field by turns. Other regions showed less interest in this area until 2017, indicating that more studies would be conducted in other regions in the future.

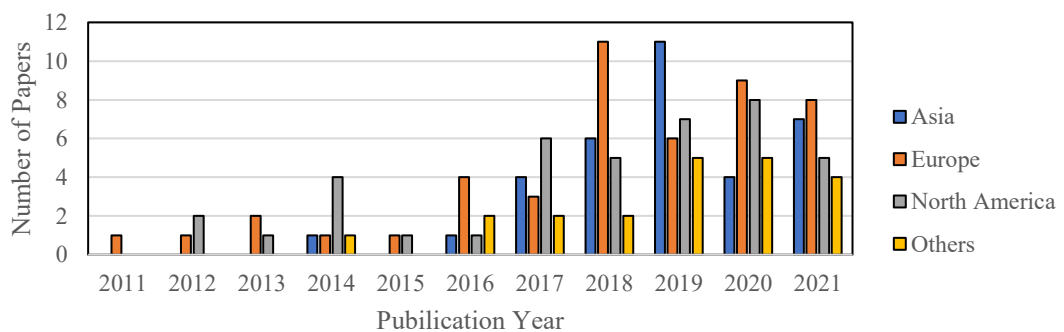


Figure 2-2 The location of case studies in the reviewed papers conducted from 2011 to 2021.

The prediction timeframe and model system are different in the identified studies, making it hard to conclude a perfect model for building occupancy prediction. However, in current studies, a typical occupancy prediction model usually consists of several procedures: data collection, occupancy prediction, and validation (as shown in Figure 2-3). Each procedure

contains various options concerning the inputs, data structure and algorithm, which require dedicated examination based on the target problem and building system. Conversely, the building performance and occupancy comfort will be impacted by the model proposed. Therefore, this paper will have the following sections: existing data gathering and sensor technology, ML techniques for developing occupancy prediction models, and model verification methodologies. The best-performing and popular predictors and ML methods will be labelled, which will help future studies construct suitable models.

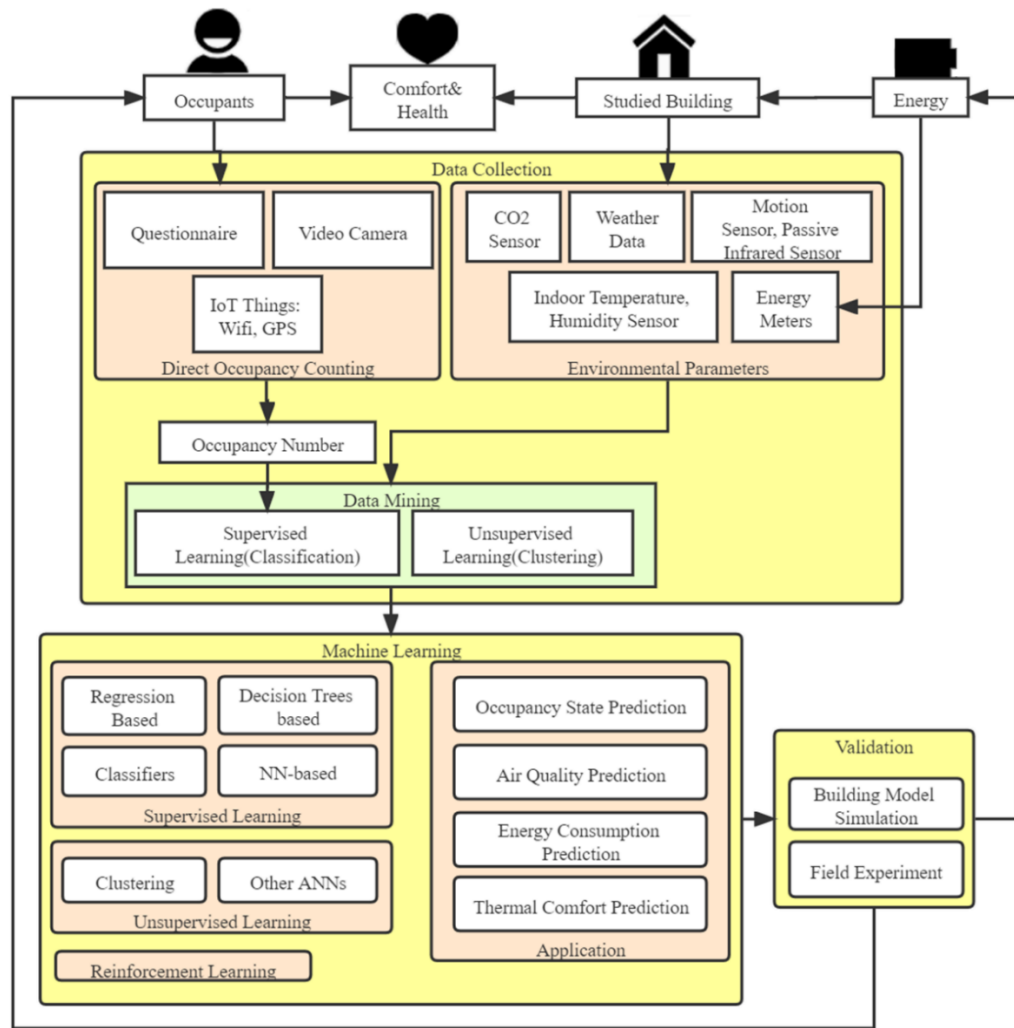


Figure 2-3 The typical procedure of occupancy prediction with machine learning, validation and applications in the built environment.

2.2 Data collection for occupancy data

2.2.1 Data collection, methods and privacy preservation

To improve the accuracy of occupancy prediction, plenty of data collection methods have recently been introduced. According to several studies, occupancy sensing can save up to 30% (Lo and Novoselac, 2010) on energy costs while improving indoor air quality (Yang and Becerik-Gerber, 2014). However, although the use of such technology is promising and provides a glimpse of future smart buildings, privacy issues have to be

addressed for wider adoption. More resolution and accurate building prediction models can be achieved by combining adequate monitoring technology of the building environment with proper HVAC or other systems monitoring.

Because the detection of occupancy status is constantly linked to privacy concerns (Nguyen and Aiello, 2013), selecting the appropriate sensor is not always simple. Based on the reviewed literature, studies are usually narrowed to academic buildings (labs or offices in universities/research institutes), which could impact the quantity and quality of data obtained, particularly when the prediction method is applied to the industry. As shown in Figure 2-4, 46% of the case studies were conducted in academic buildings. Other case study building types include office (25%), residential (16%), commercial (8%) and others such as airport terminals (Reena et al., 2018), museums (Lu et al., 2020), mosques (Aftab et al., 2017) and metro stations (Massimo et al., 2016).

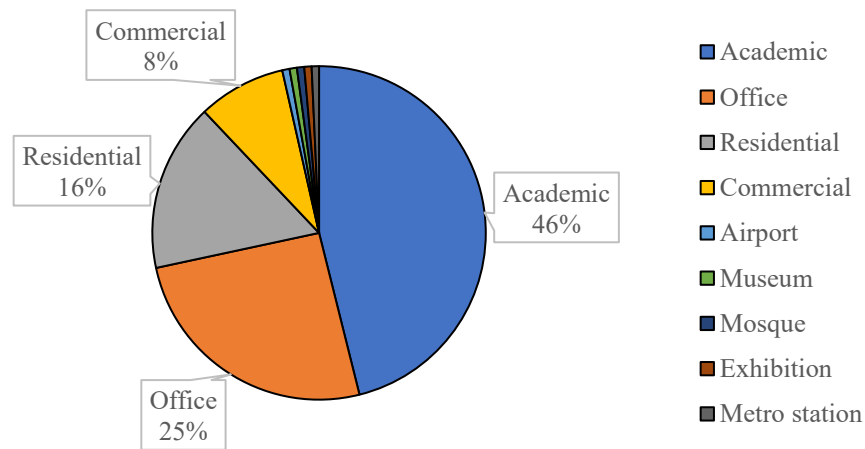


Figure 2-4 The proportion of building types in the reviewed case studies

Figure 2-5 shows the building types in case studies in different regions. Academic buildings play a dominant role in the reviewed studies in all regions because it is easier to conduct, especially when considering privacy issues. Office buildings are quite popular in all regions since the occupants are usually fixed, and not hard to get permission. In 2020, a paper conducted a case study in an office building in Stockholm,

collecting five years of data with multiple sensors installed in the building (Ferrantelli et al., 2020). However, privacy concerns may arise when such technology is applied commercially or for widespread adoption in some regions as commercial buildings are the least favoured case study type in Europe and North America. In Asia, the residential building is the least used, indicating the intense privacy concern for households in this area.

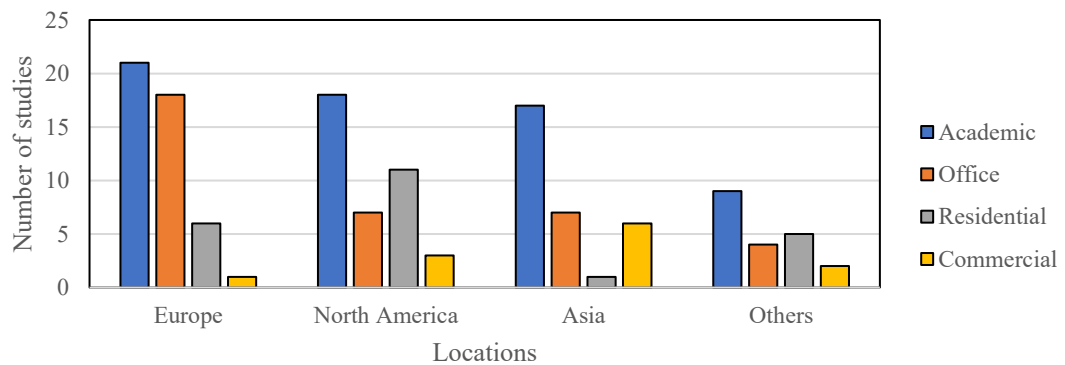


Figure 2-5 Case study building types in different regions of reviewed studies.

Privacy leakage is always a concern when choosing sensors for data collection. The key privacy risks for occupancy detection include collecting the identification and location of individuals. Masking, encryption, noise addition, anonymization of data, and scrambling of location data to avoid individual identification are all common procedures for dealing with private data. User/data anonymization is a simple solution, but it offers no protection against attackers who have direct access to the sensing database and fail to provide the room-specific information and required room identity (Alomair et al., 2010). An alternative way is to detect certain occupancy patterns in a particular zone rather than target individuals (Lee et al., 2019a). Also, occupancy location can be inferred from the occupancy data with some auxiliary information (Wang and Tague, 2014). For instance, a purposely defocused camera that creates a ‘fuzzy’ or ‘warped’ image or out-of-focus images is also a solution to room occupancy sensing (Wang et al., 2019b).

In general, the two types of data gathering methods are direct counting approaches, which directly track the occupancy number, and environmental sensors, which indirectly reveal the occupancy state. Figure 2-6 shows the connection and details of different sensors in various applications of occupancy prediction models. Temperature sensors are the most used sensors in all kinds of studies since they are easy to set up and usually pre-installed in HVAC systems or other building systems. Some sensors are only used in specific applications; for example, cameras are only found for occupancy state prediction and energy consumption prediction. Also, some sensors are more suitable for a particular application, like most studies use energy meters as sensors for energy consumption predictions. The following sections will explore the benefits and drawbacks of these sensors in terms of precision, price, ethical concerns, unresolved difficulties and future recommendations.

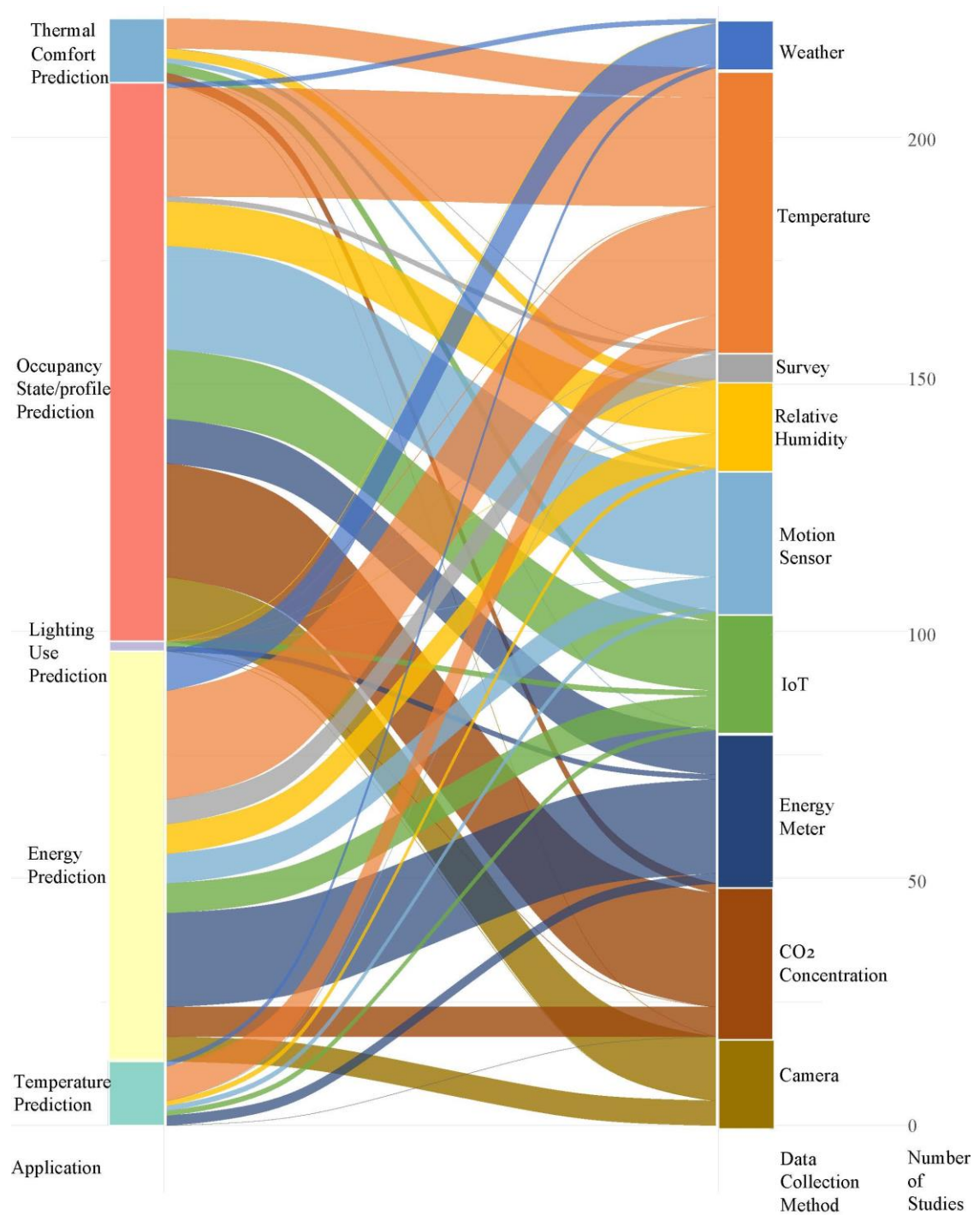


Figure 2-6 Data collection methods and their related application in the reviewed studies.

2.2.2 Direct occupancy counting sensing technology

The most straightforward way to access occupancy data or profiles that record how occupants use the facilities, or their lifestyle is to directly follow the occupants' status.




Many researchers employed questionnaires, especially data from large-scale surveys, and it is convenient for groups who share the same lifestyle, such as students on campus or residents of the same culture. A dormitory building with 200 rooms was selected as the target building, and questionnaires were sent to occupants to get their working schedules (Ding et al., 2021a). Another national survey was taken in Korea of 5240 single-person households for their daily routines (Lee et al., 2019c). Accurate occupancy data can be obtained through these large-scale surveys, and questions about their behaviour and other evaluations can be easily added to get the full picture as the research did in 2007 (Lee et al., 2019c). The mass data can show the lifestyle of a group of people. However, these surveys are usually time-consuming and require many participants from the same area and extra form-filling while participants are not always willing to cooperate.

The most accurate approach for determining the occupants' state and the number of inhabitants is camera-based occupancy detection, which is often used to offer the ground truth of occupants. An experiment in research students' office rooms with overhead cameras achieved over 80% accuracy (Wang et al., 2017) and another monitoring system with cameras was employed to examine the new proposed occupancy prediction algorithm (Li and Dong, 2017). However, most cameras were installed in the researchers' offices or specialised experimental rooms due to private intrusiveness (Chen et al., 2018).

In recent articles, wearable sensors, mobile devices, and security systems have all been used to detect occupancy (Li et al., 2020a). The Internet of Things (IoT) has opened new possibilities for occupancy detection. Wi-Fi, Bluetooth, RFID, and other technologies are examples of these strategies. Because Wi-Fi networks are common in modern buildings, they require no additional hardware or software instalments and perform well when it comes to monitoring occupancy. A Wi-Fi-based event-triggered update system for a university lecture theatre was developed in 2019 to improve detection accuracy from 77.3% to 96.8% (Lee et al., 2019b). Despite the potential for occupancy

monitoring, detection mistakes do exist, requiring extensive data cleaning methods to filter errors to acquire trustworthy occupancy data. Details of the comparisons between these data collection methods can be found in Table 2-2.

Table 2-2 Comparison and key findings between different direct occupancy counting methods in recent studies.

| Data collection method | Year | Testing environment | Study scale | Ref. | Key findings |
|---|------|----------------------------------|---------------------------|------------------------|---|
| Survey  | 2021 | Dormitory and office on a campus | 200 students and 90 staff | (Ding et al., 2021a) | ✓ Get access to the full picture of the occupancy lifestyle. ✗ Time-consuming and requires many participants |
| | 2019 | Residential houses | 5240 occupants | (Lee et al., 2019b) | |
| | 2019 | Apartments | 154 occupants | (Maljkovic, 2019) | |
| | 2007 | Residential houses | 60 occupants | (Rebaño-Edwards, 2007) | |
| Camera  | 2020 | Office | 12.4 m ² | (Tien et al., 2020c) | ✓ The most accurate method provides the ground truth. ✗ The private intrusiveness |
| | 2020 | Student centre | 1400 m ² | (Meng et al., 2020) | |
| | 2018 | Student office | 25 residents | (Wang et al., 2018b) | |
| | 2017 | Student office | 2 students | (Wang et al., 2017) | |
| | 2017 | lecture theatre | 876 m ³ | (Sultan et al., 2017) | |
| Internet of Things  | 2020 | Office | 350 employees | (Hou et al., 2020) | ✓ Low cost and requires no additional device. ✗ The detection error needs data cleaning |
| | 2019 | Residential complex | 149 rooms | (Pesic et al., 2019) | |
| | 2019 | Office | 80 employees | (Ashouri et al., 2019) | |
| | 2019 | Office | 200 m ² | (Wang et al., 2018c) | |
| | 2018 | Student office | 25 residents | (Wang et al., 2018a) | |

2.2.3 Environmental sensors for data collection

As shown in Table 2-2, most direct occupancy counting methods either cause private intrusiveness or are time-consuming. Compared to direct occupancy counting methods, environmental sensors often target a smaller group of occupants, which is partly due to the cost of sensors and the detailed data these sensors can collect. Most papers use more than one sensor to combine the data and avoid missing data. Also, when people are aware that they are being watched, they may alter their behaviour (Diaper, 1990). The ideal way of data collection would be employing existing infrastructures or simple instalments without capturing detailed personal information that concerns private intrusiveness. In most research, the case study is the researcher's own office or dwelling to avoid private intrusiveness (Tien et al., 2020b). However, the number of occupants is always limited, and the behaviour routine is usually fixed, which could make the model defective when applied to larger implementations. Therefore, some studies are conducted in public areas like shopping malls (Zeng et al., 2019) and cinemas (Arief-Ang et al., 2018b), while the sensors could miss some data with the large group of occupants.

Table 2-3 summarises some of the recent studies using environmental sensors. Many researchers use physical sensors like motion sensors to capture accurate occupancy status without being aware. 20.3% of energy-saving was achieved in a 550 m² office space with motion sensors (Peng et al., 2017), and another experiment in a smart-home testbed with a motion sensor achieved around 60% accuracy for occupancy prediction (Sama and Rahnamay-Naeini, 2016). On the other hand, motion sensors are not able to detect nearly stationary individuals, which is common in offices and during inactive time at home. Therefore, the state of occupancy can only be identified by the arrival and departure times. Also, non-intrusive sensors such as pyroelectric infrared (PIR), ultrasonic, and acoustic sensors can only be used to assess whether or not a space is occupied, not the occupants' number (Sun et al., 2014). Therefore, they are suitable for

single-occupancy rooms. For example, research conducted in a single-occupant office had a 1-hour forecast accuracy of 79% to 98% (Manna et al., 2013). However, due to the air mixing process, there were always significant delays for these sensors, especially when they were located far away from occupants.

Table 2-3 Recent studies on occupancy detection using environmental sensors.

| Ref. | Accuracy | Testing Environment | Study Scale | Data Collection Method (DCM) |
|------------------------------|--|----------------------------------|--------------------------|---|
| (Jin et al., 2021) | Up to 97.4% | Office | around 20 m ² | 3 DCM - Passive infrared sensor (PIR) sensor, an on-site survey, a camera |
| (Rueda et al., 2021) | The average accuracy of 95.8% | An apartment | - | 6 DCM - CO ₂ concentration, motion sensors, relative humidity, temperature, heating, and lighting consumption |
| (Tien et al., 2020b) | Average detection accuracy of 92.2% | Office space | 39 m ² | AI-powered camera |
| (Panchabikesan et al., 2020) | The best-adjusted R ² is 0.94 | Eight apartments | 3-bedroom apartments | 5 DCM - Motion sensors, indoor CO ₂ , indoor humidity, temperature, and the number of occupants |
| (Pigliautile et al., 2020) | Up to 84% | A house-like cubicle | 3 m x 3 m | 9 DCM - Microclimatic station air temperature, relative humidity, net-radiation, air speed, CO ₂ concentration, and illuminance level |
| (Li et al., 2020b) | Vary from 0.82-0.98 for heat consumption and 0.87-0.97 for electricity consumption | A mixed-use, university building | 7,445 m ² | 3 DCM - Outdoor temperature, and historical energy consumption data |

| | | | | |
|----------------------------|---|---------------------------|--|---|
| (Ferrantelli et al., 2020) | The error of only 5% | Office building | 8 floors, area: 19,642 m ² | 9 DCM - Water consumption, electricity load, room temperatures, ventilation devices and controllers, air pumping, indoor air quality |
| (De Bock et al., 2020) | Vary from 85.6 to 93.7% | Office | single user | 7 DCM - Motion and temperature sensors, door sensors, and pressure sensors on office chairs. |
| (Wang et al., 2019a) | The best accuracy for real-time prediction is 86% | A graduate student office | about 200 m ² with 25 residents | 3 DCM - CO ₂ concentration, relative humidity, and temperature |
| (Wu et al., 2021) | The highest R ² is 0.9594 | Office room | the floor area of 152 m ² | 4 DCM - CO ₂ concentration, temperature, relative humidity, energy consumption |
| (Peng et al., 2017) | The total control accuracy is 88.1% | Office space | 550 m ² | 5 DCM - Motion sensors, temperature sensors, relative humidity sensors, CO ₂ sensors, and HMI |
| (Sangogboye et al., 2017) | Prediction errors below 7% | A study zone | 125 m ² , 36 occupants | 6 DCM - PIR sensors, cameras, temperature sensors, CO ₂ sensors |

Therefore, environmental sensors, including CO₂-based detection, indoor temperature, relative humidity, and energy meters, are proposed. The indoor temperature sensor is the most used data collection method in the reviewed papers (as seen in Figure 2-6) because they are small and usually already available in standard HVAC systems. Since the indoor temperature is not directly linked to occupancy data, the temperature and humidity sensors are commonly combined with CO₂ sensors (Peng et al., 2017) or weather data (Marchelina et al., 2019). Also, sensors that record the indoor temperature and relative humidity are generally used to operate window openings and thermostat adjustments. These sensors, however, should be kept away from sources of heat,

humidity, and contamination (equipment, humans, and solar power) to avoid a mixture of their readings (Xilei et al., 2020).

Smart meters, which can reflect the actual electricity consumption, are also employed in many works. The energy load data is easy to collect and compare to the simulation or prediction result. Most works exhibit a significant performance gap between models and observed energy use and meters that monitor real energy consumption can be used to detect the gap and validate the influence of occupancy behaviours (Ramokone et al., 2020).

CO₂ sensors are a viable technique since they are inexpensive, tiny, non-intrusive, and non-terminal, making them a popular data collection method (Wei et al., 2019b). Since CO₂ sensors commonly exist in regular HVAC systems, no new infrastructure expenditure is needed. The method calculates the number of occupants with an equation using CO₂ concentration (Mumma, 2004), which has the main disadvantage of delayed response and possible difficulties in identifying physical parameters. As a result, when CO₂ sensors are properly installed, and details about observed rooms (room volume and airflow rate) are known, the CO₂-based method performs well, whereas the results were unreliable when the studied spaces were open and irregular, such as an open-plan or naturally ventilated office (Dey et al., 2016). To overcome these weaknesses, more accurate methods were developed including data mining algorithms.

Thermal imaging and thermal comfort voting are new contactless sensors that have demonstrated the capacity to enhance thermal comfort while affecting energy consumption. In an office room, using thermal comfort voting to obtain users' real-time reactions to the environment and then modifying the management goal settings enhances thermal comfort while saving up to 40% energy (Murakami et al., 2007). Consequently, subjective responses instead of physical parameters might be a new approach to occupancy detection that should be paid more attention to.

2.2.4 Data mining technologies

As shown in Table 2-3, in most studies, data collected from buildings has more than one kind of sensor installed. For reviewed papers in this article, the most widely used method is the combination of indoor temperature sensors and CO₂ sensors (Khalil et al., 2021, Yuan et al., 2020, Wang et al., 2019a). However, raw data might have a variety of issues, such as missing information or sudden swings if one or more sensors are disrupted. Also, sensor readings could conflict with each other, and sometimes, the reading in sensors will not change much, so it provides no valuable information.

To solve these problems above, data mining technologies have been introduced by many researchers. For example, missing data were replaced with interpolated data, and nonsensical data was either removed or reset to the sensor's initial values using the "data cleaning" method (Yu et al., 2011). Extraction of the mean, standard deviation, mean absolute deviation, first, second, and third-order differences and even simple moving averages are used as post-processing procedures for collected original data. For data mining, most researchers use supervised algorithms like the SVM (Support Vector Machine) and the Decision Tree to categorize samples based on a target variable (Das et al., 2019). Unsupervised learning techniques, such as hierarchical clustering and k-means, have recently been adopted in studies to organize data into clusters based on the characteristics of all variables without any target variable (Killian and Kozek, 2019). With the trend of multiple sensors, it is hard to confirm an occupancy dataset structure in advance. Therefore, using cluster algorithms is becoming a standard step before sending data to machine learning training.

2.3 Machine learning algorithms and their applications

Supervised learning, unsupervised learning, and reinforcement learning are the three most typical machine learning approaches used in occupancy prediction (Mohri et al.,

2012). Supervised learning models include decision trees (Koklu and Tutuncu, 2019) (such as the gradient boosting tree), classifiers (such as the Bayes classifier, kNN, and support vector machine), and neural network-based models (Kim et al., 2020) (such as the feedforward backpropagation network and cascade correlation). Furthermore, these models can be classified as linear or nonlinear based on the data structure. Linear methods are used when the responding and prediction data are linearly linked or converted into a linear relation. With the dramatically increasing of variates, data transformation techniques like normalization process, log conversion, and ranking transformation might be utilized (Apostolo et al., 2020). In the majority of circumstances, linear models are easy to create and use, and they are frequently used as the first model. Other nonlinear models can be employed more effectively if the data are unlikely to be linearly connected.

Unsupervised learning methods reduce, summarize, and synthesize data using unlabelled training data (Mohri et al., 2012). Unsupervised learning algorithms include cluster analysis learning, like principal component analysis and parametric analysis, and various ANNs (e.g., autoencoder neural network and self-organising map) (Liang et al., 2016). Because occupants behave in a stochastic manner impacted by a variety of parameters, the majority of which are immeasurable and unpredictable, it's critical to figure out which inputs are the greatest influencers and only add those that significantly increase behaviour. As a result, while unsupervised learning cannot generate prediction for a new dataset, it can contribute to the comprehension of the data's character, allowing for the selection of supervised models for prediction (Amel et al., 2018). Since there is no output in unsupervised methods, data linearity is not an issue. Similarly, with reinforcement learning, a direct match of input and output does not exist, and it can only estimate how well the output is.

2.3.1 The trends of machine learning and deep learning

In general, there is a rise in machine learning applications because of the availability of building automation systems, smart systems and IoT platforms, which increases the quantity of data available as discussed before (Zantalis et al., 2019). The great volume of data requires advanced techniques to analyse them which conventional models cannot handle properly. In addition, most behaviours are influenced by several contextual elements, the best way to mimic them is to either integrate all the parameters in one equation or address the factors that influence behaviour separately, allowing them to be split into various formulae. Therefore, powerful methods like deep learning which is suitable for big-data and computationally intense processes have been introduced in recent years.

As can be seen in Figure 2-7, the neural network-based algorithm (which occupied more than 40% of reviewed papers after 2018) is the most popular method in building machine learning prediction. Particularly, deep learning with a large number of hidden layers that compose the neural network showed good capacity in image pattern recognition, speech recognition and synthesis, etc. which also indicated possible future development in occupancy prediction models.

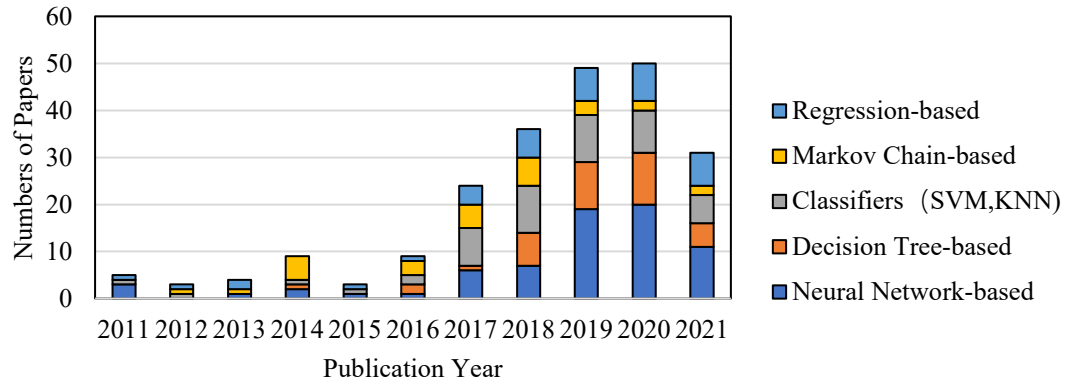


Figure 2-7 Summary of the reviewed studies from 2011 to 2021 using machine learning algorithms.

The popularity of the neural-network-based algorithm indicated that deep learning is making major advances as typical machine-learning techniques were narrowed in the ability to deal with data in the natural form (LeCun et al., 2015). Deep learning uses graph technologies and neuron transformations to obtain multilayer learning models and automatically learns the data. The most widely used deep learning models are Convolutional Neural Network (CNN) (Arvidsson et al., 2021) and Recurrent Neural Networks (RNN) (Kim et al., 2018), which are also popular in building occupancy prediction. Also, the development of deep learning algorithms provides advancement in building automation systems as it can convert the data at one level (starting with the natural data) into a depiction at a slightly more abstract level. In 2021, a smart Oracle-based building management system was proposed that auto-learns occupancy patterns and leverages spatial organization to deliver actionable insights on energy savings (Mitra et al., 2021).

2.3.2 Occupancy prediction

Occupancy prediction, in general, draws the most attention in the reviewed papers until 2020, which is since the variation of occupants' interactions is regarded as the foundation of the uncertainty in building models. One of the key parameters an

occupancy prediction model should consider is the occupancy level. In 2011, a study separated occupancy prediction levels into three major factors: temporal, spatial and occupancy state resolution (Melfi et al., 2011). The precision with which the timing of events is modelled is referred to as temporal resolution. The precision of the physical scale is defined by spatial resolution (e.g., a building or a zone of the model target). The model's occupancy resolution refers to how it specifies individuals.

For temporal resolution, one of the classifications divided occupancy prediction models into three categories: real-time recognition, future time-step predicting, and occupancy profile modelling (Xilei et al., 2020). These approaches either estimate the number of occupants, determine whether they exist in a particular area, or generalise a few occupancy profiles based on previous occupancy patterns. In most occupancy prediction models, the monitor period has ranged from a day to multiple years, and the time they mean to predict varies from several seconds to more than a year. A study showed 61.5% and 43.6% accuracy for building predictions of more than a year and one hour, respectively (Kuster et al., 2017). The short-term prediction has a direct application for quick occupancy demand response and suits the needs of the industry. However, the seasonal effect of occupant behaviours requires a full year of monitoring is more reliable, especially in specific cases, such as simulating academic buildings' holiday schedules regarding energy consumption (Li et al., 2020b).

In the reviewed studies, the regression-based method based on Random Forest is the most used method for occupancy prediction. According to a study, the regression model is primarily used for long-term forecasting, while ANN is mainly used for short-term forecasting (Kuster et al., 2017). Different methods should be employed for different types of occupancy state prediction. For example, ANN with long short-term memory (LSTM) architecture is the most commonly used and suitable method for time series prediction (Mandic and Chambers, 2001). A study found that Random Forest is the most suited classifier (Haidar et al., 2019), with at least 90.53% accuracy, after training data

with five different machine learning classifiers (Random Forest, Decision Tree Classifier, Extra-Trees, Gaussian Naive Bayes, and Multi-layer Perceptron). Another study achieved 97.27% to 98.90% accuracy in an indoor office by employing several Deep Neural Networks (DNNs) (Metwaly et al., 2019a). The method's accuracy also depends on the type of data collected. For example, the SVM and k-NN models have lower counting errors when using Wi-Fi data, whereas the ANN model is more accurate when using fused data (Wang et al., 2018a).

Based on the reviewed papers, many studies focus on detecting the occupancy information, including the occupancy presence, number and location in space, zone or building. However, there are limited studies on the detection of occupancy activities, for example, movement in space (Tien et al., 2020c), opening/closing of windows (Tien et al., 2022), adjustment of HVAC, and use of equipment and appliances. Furthermore, significant attention of the existing literature is focused on the performance of developed algorithms, such as their speed and accuracy. Details of different kinds of occupancy prediction are listed in Table 2-4.

Table 2-4 The information about studies using various algorithms in occupancy state/number/activities prediction.

| Prediction Classification | Ref. | Year | Sensor | Algorithm | Test Environment | Accuracy |
|-----------------------------|---------------------------|------|---|---|---|----------------------|
| Occupancy State Prediction | (Vaňuš et al., 2019) | 2019 | Relative humidity, temperature, and CO ₂ | Linear Regression, Neural Networks, and Random Tree | A laboratory | Higher than 90% |
| | (Arief-Ang et al., 2018a) | 2018 | CO ₂ data and indoor human occupancy | seasonal-trend decomposition (STD) | An academic office and a cinema theatre | An average of 94.68% |
| Occupancy Number Prediction | (Apostolo et | 2021 | 28 Wi-Fi Apps | Multilayer Perceptron ANN | 5 floors of classrooms | RMSPE of 0.29 |

| | | | | | | |
|-------------------------------------|--------------------------------------|------|--|---|--|--------------------------------------|
| | al., 2021) | | | | | |
| | (Salim i et al., 2019) | 2019 | Real-Time Locating System | inhomogeneous Markov chain | A research laboratory | 86% on average |
| | (Wang et al., 2018c) | 2019 | Wi-Fi probes and indoor air temperature , relative humidity, and airflow rate | Gradient tree boosting, Random forests, AdaBoost | A large office room, 200 m ² | Reached 72.7% |
| | (Kim et al., 2019) | 2019 | Camera and motion sensor | RNN with LSTM units | An exhibition | Best RMSE of 10.31 |
| Occupancy Activity Prediction | (Tien et al., 2020b) | 2021 | Camera | CNN | Office space, 39 m ² | Average accuracy 92.2% |
| | (Lu et al., 2020) | 2020 | Social networks | Random Forest and XGBoost | A public museum | RMSE within 30% |
| | (Huch uk and Sanner , 2019) | 2019 | Temperatur e sensor and PIR sensor | Markov model (MM), HMM, and RNN | Single-family homes | Under 0.80 average accuracy |

Limited works focused on evaluating the impact of the detection technique on the performance of the building and HVAC systems. For example, a study proposed a vision-based approach for detecting and recognising the occupants' activities within building space (Tien et al., 2020c). Unlike previous works which focused on occupancy levels, the study used the data to predict the indoor heat gains from the occupants with varying activity levels. Such information would be useful for HVAC controls to adapt and make a timely response to dynamic changes in occupancy activities. A recent work used the same detection approach to detect how the occupants interact with the equipment or appliances such as computers (Tien et al., 2022). Similarly, the proposed approach can predict the internal gains from the facilities operated, contributing to the indoor heat gains.

The exploration of vision-based methods for occupancy detection and prediction in buildings has gained considerable attention in recent years. Table 2-5 outlines examples

of recent research efforts in this area, highlighting a variety of algorithms and data collection techniques utilised over the years. For instance, Yang et al. (Yang et al., 2022) employed a framework that utilises CNN-based density estimation methods to fuse image information from surveillance videos to obtain accurate and high spatial-temporal resolution indoor occupancy information. This framework trains an ML-based ensemble model to predict occupancy schedules based on the occupancy information extracted from the images, achieving 95.67% accuracy in high-density environments.

Table 2-5 Examples of research work on occupancy detection and prediction using the vision-based method

| Ref. | Year | Data Collection | Test Building | Participants number | Result | Algorithm |
|----------------------------|------|------------------------------------|---------------|---------------------|-------------------------------------|--|
| (Zou et al., 2017) | 2017 | Surveillance camera | Office | 12 | The number of occupants | CNN, SVM, and K-means |
| (Callemein et al., 2019) | 2019 | Omnidirectional camera | Office | 4 | The number of occupants | YOLOv2 |
| (Tien et al., 2020d) | 2020 | Camera | Office | 1 | The number and activity of occupant | CNN |
| (Sun et al., 2022b) | 2021 | Entrance video and interior camera | Office | 11 | The number of occupants | GMM, CNN-based FCHD, Kalman filter, OFH |
| (Choi et al., 2021) | 2021 | Internet protocol camera | Office | 10 | The number of occupants | YOLOv5 |
| (Tien et al., 2022) | 2021 | Camera | Classroom | 2 | The number of occupants | Faster R-CNN with Inception V2 |
| (Sun et al., 2022a) | 2022 | Cameras | Office | 11 | Head and occupancy detection | YOLOX, Deepsort, dynamic Bayesian fusion (DBF) |
| (Gursel Dino et al., 2022) | 2022 | Cameras in large indoor space | Classroom | More than 100 | Head and occupancy detection | YOLOv3 |
| (Choi et al., 2022) | 2022 | Internet protocol camera | Office | 6 | The number of occupants | YOLOv4 |
| (Wei et al., 2022a) | 2022 | Camera | Classroom | 7 | Occupancy count and | Faster R-CNN with Inception V2 |

| | | | | | | |
|---------------------|------|---------------------|-----------|---------------|---------------------------------------|--|
| | | | | | activity profiles | |
| (Wei et al., 2022b) | 2022 | Camera | Classroom | 3 | Occupancy count and activity profiles | Faster R-CNN with Inception V2 |
| (Yang et al., 2022) | 2023 | Surveillance videos | Classroom | More than 100 | Occupancy count and schedules | CNN-based density estimation method and ensemble model |

Sun et al. (Sun et al., 2022a) proposed a three-level fusion framework based on YOLOX for indoor occupancy estimation in a University office space, which achieved a prediction accuracy of up to 99%. Furthermore, Choi et al. (Choi et al., 2022) employed a deep learning model based on YOLOv4 for occupancy counting in small and medium-sized offices, demonstrating high performance (root mean square error (RMSE): 0.883), broad applicability and cost-effectiveness of the method.

Expanding on these advancements, Sun et al. (Sun et al., 2022b) proposed a four-step system combining motion detection and static estimation. This approach filters non-occupied frames, detects entrance and exit events, and uses a Fully Convolutional Head Detector (FCHD). The results are fused using Kalman filtering and Occupancy Frequency Histogram (OFH), achieving 97.8% accuracy. This fusion method effectively addresses common issues like occlusions and cumulative errors, enhancing occupancy estimation in diverse environments. These studies (Wei et al., 2022a, Tien et al., 2020d) have shown that with accurate real-time occupancy data in building management systems, HVAC systems can be optimised to match actual occupancy patterns, which leads to improved energy efficiency and significant energy savings. For instance, the study (Han et al., 2024) reduced the total energy consumption of fan coil units by 18.43%, 8.71%, and 18.97% in different cities by using the occupancy estimation framework.

Some studies have demonstrated the capability to identify not just the presence of occupants but also specific activities and behaviours such as using equipment (Wei et al., 2022a) or opening windows. Such activities can influence the heat gains or heat loss in

buildings and consequently influence the operation of HVAC systems. For instance, Wei et al. (Wei et al., 2022b) utilised Faster R-CNN models to detect the number of occupants and their specific activities such as walking, sitting, and standing, achieving an accuracy of up to 88.5% for activity recognition. However, the accuracy of detecting specific activities is generally lower compared to simply detecting the number of occupants, which can achieve higher accuracy, as evidenced by the 98.9% accuracy for occupancy counting in the same study.

In another study, Wei et al. (Callemein et al., 2019) introduced a real-time occupancy and equipment usage detection approach using Faster R-CNN for demand-driven controls. The approach achieved 93.60% for occupancy activity detection but lower accuracy for equipment detection (78.39%). Occupancy activity detection requires the detection of the entire body of the occupants, which can be more challenging than methods that employ head counting or detection. Similarly, Tien et al. (Tien et al., 2021) employed Faster R-CNN to detect real-time window conditions, achieving a detection accuracy of 97.29% in tests conducted in a case study building. Building on this, the study (Tien et al., 2022) developed a real-time occupancy and window operation detection, which achieved 85.63% for occupancy activity detection and 92.2% for window operation detection. These studies highlight the potential for reducing energy loss and optimising HVAC systems by incorporating real-time detection of window operations.

The use of cameras for occupancy detection raises privacy issues, as continuous monitoring can be intrusive. A small survey conducted by (Choi et al., 2021) suggests that people preferred occupancy counting techniques that automatically extract and delete images without human intervention, rather than methods that use video encryption or blur occupants. Callemein et al. (Callemein et al., 2019) addressed privacy concerns by using a low-resolution omnidirectional camera that maintains privacy while still providing accurate occupancy counts. Using YOLOv2, they combined spatial and temporal image data to improve detection performance even with extremely low-resolution images.

Many of the studies highlighted above were conducted with a limited number of participants in small to medium-sized offices and classrooms, indicating a potential area for further exploration and validation of these vision-based methods in more populated building spaces. A potential limitation in larger scenes is that the increased distance between the camera and occupants results in lower-resolution images (Gao et al., 2020a), making it difficult for detection algorithms to accurately identify and count individuals. Furthermore, the complexity of indoor environments, such as open-plan offices and classrooms, further poses a challenge due to the presence of obstacles like furniture and equipment, which can hinder accurate person identification (Zhang et al., 2016).

To address these issues, some studies have explored the use of multiple cameras for localising and counting individuals (Choi et al., 2021). For instance, one study (Maddalena et al., 2014) introduced a 3D self-organising neural network approach utilising multiple cameras to tackle occlusions and visibility issues common in crowded and cluttered scenes. The use of multiple cameras can provide different viewpoints and help overcome occlusion problems by covering blind spots, but these methods often necessitate calibrated and synchronised cameras, introducing a layer of complexity and computational demand.

Dino et al. (Gursel Dino et al., 2022) investigated vision-based methods for estimating the number of occupants using multiple video cameras. Their hybrid approach combined instantaneously counting people in a scene with incrementally counting those entering or exiting a room. Tested in a large, crowded, and occluded classroom with over 100 occupants, it showed high predictive capacity. However, the study focused on head detection and counting occupants, without considering specific activities or the impact on building energy performance. It also noted high computational costs and significant infrastructure expenses for multiple cameras and continuous internet connections. This underscores the need for more efficient algorithms and comprehensive privacy-preserving solutions. Furthermore, Yang et al. (Yang et al., 2022) proposed a CNN-ML framework

for crowd counting and prediction in high-density public buildings, achieving 95.67% recognition and 83.12% all-day prediction accuracy. However, detection accuracy decreases in dense scenes, and the method has high computational costs and requires extensive manual labelling. Future research should optimise algorithm efficiency and reduce manual labelling.

In contrast, another study (Callemein et al., 2019) employed an omnidirectional or wide field-of-view camera mounted in the ceiling, which captures a single 360-degree image without the need for camera repositioning. This method simplifies the setup by using a single camera to cover an area, reducing the complexity and cost associated with multiple cameras. However, it introduces challenges related to image distortion and lower resolution at the edges of the captured image. The study achieved favourable results but required retraining the detector with similar omnidirectional images to account for the distortion effects. While multi-camera systems improve accuracy in complex environments, they pose challenges in cost and complexity. In contrast, single-view camera systems are simpler and more cost-effective. This research will focus on single-view camera systems evaluating their potential for accurate and efficient occupancy detection in building environments.

2.3.3 Indoor air quality (IAQ) prediction

IAQ has long been an important topic for the health and wellbeing of the occupants in buildings. The previous sections have highlighted the importance of a holistic approach to deal with these challenges adequately. Traditionally, mechanistic IAQ models have been utilized, and the link between inputs and outputs has been based on mechanisms (Yang et al., 2014). However, mechanistic IAQ models do not include the interactions between the occupants and the indoor environment and the differences between individuals, which can impact energy consumption and building performance. The operation of HVAC systems affects both comfort and IAQ. Hence in some studies, IAQ

prediction is combined with thermal comfort prediction and considered as part of the overall occupant's comfort parameter (Goyal et al., 2012). Therefore, these models, especially ML models, which consider occupancy interaction and building performance, are increasingly being employed in recent research.

One of the most crucial issues in IAQ prediction is finding the right input to achieve a reliable prediction. Since the model is data-driven, it is important to identify the key variables inputs. Many environmental indexes are used to determine the relationship between occupants' feelings about IAQ, such as door/window opening behaviour, temperature, relative humidity, CO₂ concentration, solar radiation, rainfall, wind speed, noise, illumination, and so on (Kallio et al., 2021). Therefore, in IAQ prediction models, normally, one or more driving factors are used for prediction (Kamel et al., 2020). The inputs may have an indirect and unexpected impact on the behaviour (Killian et al., 2018), therefore an over-fitted model that has many inputs is often conducted. Many research used data mining approaches such as stepwise regression, principal component analysis (Das et al., 2014), and partial least squares (Kim et al., 2009) to uncover the driving components before developing the models.

The algorithm selection is often related to the data structure and collection method. However, IAQ is related to a lot of environmental indexes, as stated before, which can be recorded by various kinds of sensors and parameters, it is hard to recommend a specific kind of algorithm without analysing the detail of the model. One review summarised the popular algorithms, for example, ANN, linear regression models, and Decision Trees developed for predicting different factors of IAQ, but cannot recommend the optimal method and suggested a test and compares different models before choosing the most suitable model (Wei et al., 2019a).

Most IAQ models are employed to improve the occupants' overall comfort or lower the concentration of indoor air pollutants. For example, a study tested a control model

of a filter for indoor CO₂ decreasing in a sports centre while using fuzzy inference to reduce the indoor CO₂ concentration (Omarov et al., 2021). Another research found Multilayer Perceptron (MLP) follows the pattern of CO₂ changes more quickly and with higher accuracy compared to other algorithms (Support Vector Machine (SVM), AdaBoost (AdB), Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LR)). It reduced 51.4% of energy consumption in the total energy usage (Razban and Taheri, 2021). Other environmental indexes like PM2.5 concentration can also be predicted by neural networks (i.e., RNN, LSTM, and gated RNNs) (Loy-Benitez et al., 2019).

2.3.4 Thermal comfort prediction

The number of thermal comfort prediction studies and approaches using the ML methods is limited compared to occupancy and energy consumption prediction, as in Figure 2-1. In the existing literature, thermal comfort is typically assessed by the PMV model based on extensive laboratory tests, which ignore individual comfort (Fanger, 1970a) and, in some cases, do not provide satisfaction for all occupants (Cheung et al., 2019a). Therefore, most existing literature uses the ML approach to forecast thermal comfort and consider all occupants as a whole, disregarding data acquired from separate occupants (Chai et al., 2020). In this scenario, individual occupant diversity was lost, and occupants were modelled as an "average group," which is a statistical construction rather than an actual person (Goyal et al., 2012). It's worth noting that occupant comfort differs according to one's age, gender, background, and other personal characteristics. Therefore, individual comfort is becoming more popular, and personal comfort models based on data from individual occupant comfort surveys are being developed (Issaraviriyakul et al., 2021).

A recent study used two different machine learning algorithms to analyse a combination of inputs, including an individual comfort system, body temperatures, timing, and

environmental parameters. Personal comfort models achieved the best accuracy across all examined methodologies and participants, according to their findings (Katić et al., 2020). With the advancement of the Internet of Things, it is becoming more convenient to collect physiological data using a range of sensors (wearable or non-wearable devices). They can forecast thermal experience or satisfaction based on users' physiological data, such as employing wearable devices to monitor skin temperature, heart rate, blood pressure, and other physiological parameters at various human body positions (such as wrist, face, back and legs) (Chai et al., 2020). Therefore, these sensors show potential for the future development of thermal comfort prediction.

In 2012, the research employed a PMV control model with an RNN network and branch-bound boost to the HVAC system (Ferreira et al., 2012). Another study looked at the effectiveness of an ANN-based adaptive PMV control algorithm in a residential house and discovered that it was more effective than non-adaptive algorithms for improving control and disturbance reaction (Moon, 2012). Meanwhile, since two behaviours can achieve the same goal and thermal comfort often links to several behaviours, ensemble models are likely to be introduced in comfort prediction models. A paper using the machine learning approach Bagging, using a multilayer perception network (MLPN) as a learning algorithm, outperformed traditional ANN and SVM methodologies (Wu et al., 2018).

The prediction model of thermal comfort is directly linked to the occupants' satisfaction with the indoor environment. With new ML models and data collection methods, the performance gaps will be reduced. Improved models could be linked to a real-time environmental control system to improve building management without sacrificing occupant comfort. For example, as shown in Figure 2-8, the environmental information obtained can be used to provide data for the prediction of thermal comfort in real-time, which can be used to adjust the operation of the HVCA system. The occupancy data, such as the occupant's number and metabolic/activity level, can estimate the indoor CO₂

level and minimum ventilation level. Similarly, the thermal comfort prediction model can also use the occupancy number and activity level. Such information can be used to optimise the HVAC operation while also minimising the energy demand.

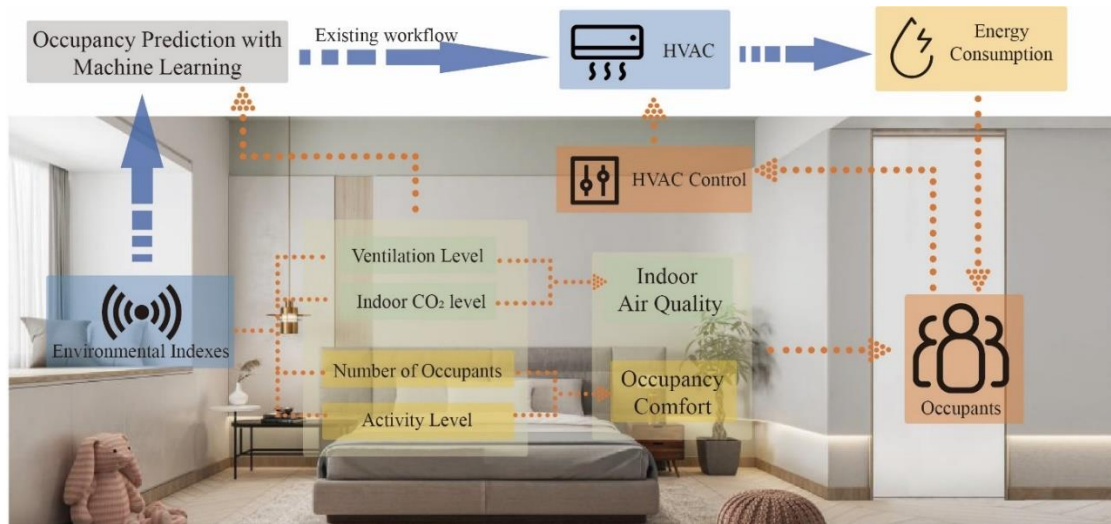


Figure 2-8 The existing workflow of IAQ and thermal comfort prediction and the potential improvement.

2.3.5 Energy consumption prediction

Prediction of energy usage in buildings is becoming increasingly important, however, it is influenced by interrelated physical, operational, and behavioural factors such as building material, building schedule, and occupant behaviour (Chari and Christodoulou, 2017). In most cases, physics-based building energy simulation tools (white-box models) such as DOE-2 and Energy-Plus are often used (Fumo et al., 2010). However, these tools are limited for energy analysis since they do not contain uncertain factors like occupancy behaviour, impacting annual energy consumption up to 75% for residential buildings and 150% for commercial buildings (Clevenger et al., 2014). As a result, many researchers use the data-driven method (black-box models) to forecast energy use and analyse the effects of energy-saving initiatives like energy-retrofit strategies and renewable energy technology (Wei et al., 2018). Meanwhile, other researchers use the

output of occupancy prediction to generate an occupancy profile as input for physics-based simulation tools to calculate the energy use result (grey-box model).

Existing machine learning-based models, on the other hand, do not adequately account for occupant behaviour. They either ignore occupancy behaviour entirely or deal with it in a limited way, such as merely examining building operation schedules (Massana et al., 2016) or simplifying the occupancy model as occupancy rate (Truong et al., 2021). In addition, with the new development of data collection methods, models that target specific occupants will be proposed. A model simulating energy consumption on the personal level and considering the gender difference was proposed in (Lee et al., 2019b) and concluded that females tend to use more energy than males.

Although HVAC is usually required to provide comfortable, productive, and healthy surroundings, it also uses a large amount of energy (Fan et al., 2017). However, occupants have many adaptive opportunities and other energy-relevant behaviours to minimise consumption. Furthermore, two behaviours can achieve the same goal, for instance, adding more clothing and turning on or adjusting a heater can both lead to warming a person, but at different levels of efficiency, price, and energy intensiveness. Most machine learning models of energy consumption only evaluate and discuss a single behaviour without considering their correlated relationship. It could be due to the ML algorithm requirements for the data structure and simplifying the model. Therefore, choosing suitable inputs and model structures is critical for the prediction method and affects accuracy and performance.

The most often utilized methods for building energy estimates using historical data are regression and ANN models (Ahmad et al., 2018). The performance of different data-driven models may differ from residential, commercial, and office buildings when picking the best strategy for a certain case. Most researchers would use a trial-and-error method to find the best model performer for a certain structure instead of assuming a

universal model and applying it to all building types. In general, ANN prefers environmental, time index inputs (Kuster et al., 2017). Ensemble models, which combine numerous models due to the nature of energy use in buildings, are more likely to produce accurate predictions than single models (Wang et al., 2018d). A few studies achieved better results with the ensemble techniques than the single method. For example, the performance of three ANN models – Feed Forward Neural Network (FFNN), radial basis function network (RBFN), and adaptive neuro-fuzzy interference system (ANFIS) – was compared to the ensemble of these three models, and the ensemble model produced the best accurate prediction results (Jovanović et al., 2015).

One major challenge to the machine learning model is the large number of algorithms available, making it difficult to determine which one should be used for a given task. The type of data provided determines the learning methods. Statistical models are classified as linear or nonlinear based on whether they are used to solve linear or nonlinear problems. After appropriate data transformations, nonlinear issues can be turned into linear ones. Aside from the differences, one model may involve multiple learning algorithms, with their own set of strengths and disadvantages, making it even more difficult to choose the best method. Making several assumptions and testing various approaches is a frequent solution. A more comprehensive estimation can be obtained by training various models and combining prediction outcomes. Consequently, it is vital to summarize the data for various applications to assist researchers in developing better prediction models. A list of popular machine learning algorithms for different applications in the existing literature is made in Table 2-6.

Table 2-6. Summary of the commonly used machine learning algorithms for different applications

| Application | Algorithm | Suitable Cases | Accuracy | Ref. |
|-------------|-----------------------|---|-----------------|----------------------|
| | Decision tree and HMM | The decision tree is suitable for current | 86.2% and 93.2% | (Ryu and Moon, 2016) |

| | | | | |
|--------------------------------------|--|---|--|--------------------------------|
| Occupancy State Prediction | | state detection and HMM for future state | | |
| | CNN | Good with images | 89.39% | (Tien et al., 2020a) |
| | DNN | Suitable for resource-constrained devices used in IoT-based applications | Ranging from 97.27% to 98.90%. | (Wang et al., 2018a) |
| Indoor Air Quality Prediction | LSTM | Outperform other algorithms with real-time collected data | 96% | (Hitimana et al., 2021) |
| | Markov model and ANN | Markov model for comfort assessment and ANN for CO ₂ predictions | $R^2 = 0.92$. | (Tagliabue et al., 2021) |
| | SVM, AdB, RF, GB, LR, and MLP | MLP outperformed for CO ₂ forecasting | The best RMSE for MLP is 33.78 | (Razban and Taheri, 2021) |
| Energy Consumption Prediction | k-means cluster | Better fitting for time series with less mobility of occupants or the rooms with larger capacity | 15 % error | (Ding et al., 2021a) |
| | ANN four Back-propagation neural network | Levenberg–Marquardt Back-propagation has better performance in forecasting electricity consumption. | The error rate is 1.07–2.23% | (Kim et al., 2020) |
| | SVR, LMSR, KNN and NB | Regression models fit for modelling daily electricity and heat demand | varies from 0.82–0.98 for heat consumption and 0.87–0.97 for electricity consumption | (Li et al., 2020b) |
| | LSTM and NNARX and MLP | LSTM models reduce prediction error by 50%. | the error is under 0.35 | (Mtibaa et al., 2020) |
| Thermal Comfort Prediction | SVC and ANN | Suitable for single-room residences with the phone application | above 95% | (Issaraviriyakul et al., 2021) |
| | ANNs and SVM, PMV, aPMV, and ePMV | ANNs model is effective in naturally | ANNs model had the highest R | (Chai et al., 2020) |

| ventilated residential buildings | | (0.6984) and R ² (0.4872) values | |
|--|--|---|----------------------------|
| Linear Discriminant Analysis (LDA), KNN, DT, NB, SVM, and RF classifiers | Could be combined with the real-time control system | up to 84% | (Pigliautile et al., 2020) |
| LSTM | Can accurately forecast overheating conditions throughout the year | over 95% | (Yuan et al., 2020) |

Therefore, new prediction methods that distinguish different types of activities and the personnel management system are required for future energy consumption models and fill the research gap. Like the methods discussed in the earlier sections, future energy models could benefit from more advanced occupancy data collection methods or integrated sensor systems, which can better capture the dynamic variations and make the necessary adjustments to the HVAC system.

2.4 Validation of the prediction models: case study and time series

Most studies include a validation stage or process after obtaining the results, which evaluates the proposed model's accuracy and applicability. The leave-one-out cross-validation approach is the most common validation method. The entire data set is usually separated into three sections: training stage, verification, and testing. The majority of the data is normally used for training (more than 70%), while the rest is used for testing and model validation (Arief-Ang et al., 2018a). The result from machine learning methods will be compared with the validation data collected to evaluate the method's accuracy.

In the reviewed studies, as shown in Figure 2-10, most research conducted field experiments in existing buildings or testbeds to test and validate the proposed method, while others used simulation-based investigations. Using historical occupancy data or other data collected as the input, the prediction accuracy can be up to 95% (Dey et al.,

2016). For experimental studies, the implementation scale in reviewed studies varies from a small testbed (Gilani and Gunay, 2018) to the whole building (Ding et al., 2021a). Many energy-related experiments are conducted in a whole building, while most occupancy prediction models use selected rooms inside a building for the case study (Figure 2-9). Some research separates the testbed into zones to compare different methods (Rahaman et al., 2019), while others define a small area as a testbed to check the prediction method (Pigliautile et al., 2020). The selection of implementation scale is often related to experiment design, and the challenges researchers faced ranged from communication issues with facility managers to equipment (De Bock et al., 2020) and sensor malfunction, which should be considered before conducting similar experiments (Gilani and Gunay, 2018).

Some of the studies use public occupancy datasets to test the prediction models they proposed. For example, one research employed the ASHRAE Global Thermal Comfort Database with data from 52 field studies conducted in 160 buildings around the world (Földváry Ličina et al., 2018). This database is also used in another project to study the subjective metrics used for the assessment of the occupants' thermal experience (Wang et al., 2020). Another example is the American time use survey (ATUS) conducted by the U.S. Bureau of Labor Statistics as an annual survey to record the respondent's activities and locations on a regular day (Statistics, 2009). Another dataset conducted in 2015 in Berkeley, California includes whole-building and end-use energy consumption, HVAC system operating conditions, indoor and outdoor environmental parameters, as well as occupant counts (Luo et al., 2022). With the awareness of the importance of occupancy behaviour, there will be more datasets available in the future and validated by the scientific community.

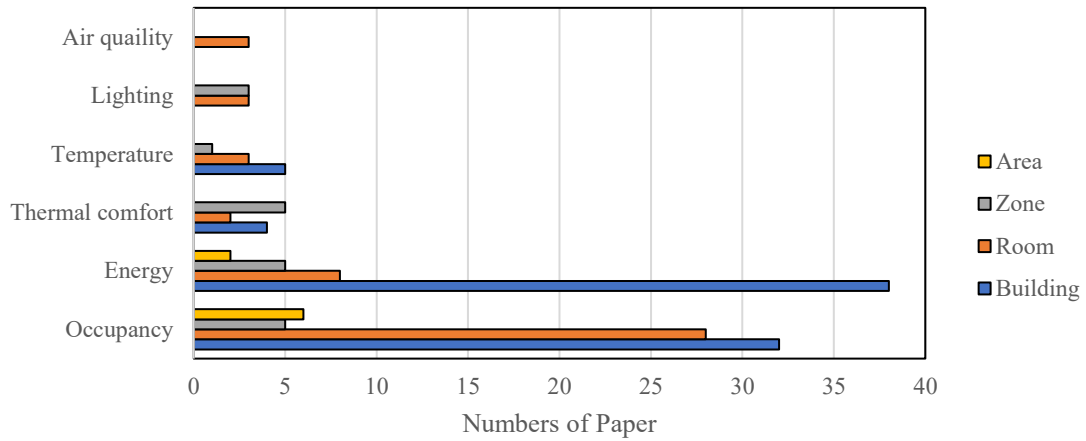


Figure 2-9 The implementation scale of different applications of prediction models in reviewed studies.

Also, the time series they meant to predict in Figure 2-10, short-term, long-term, and 24-hour predictions each contribute about one-third of reviewed papers for all regions. Short-term predictions are more common in North America and Europe, while long-term predictions are more common in Asia. This could be due to the sensor chosen and the prediction method design difference and most of the short-term predictions are usually tested before the longer version. The time series in different regions is shown in Figure 2-10, as the red columns indicate the time length in implementations.

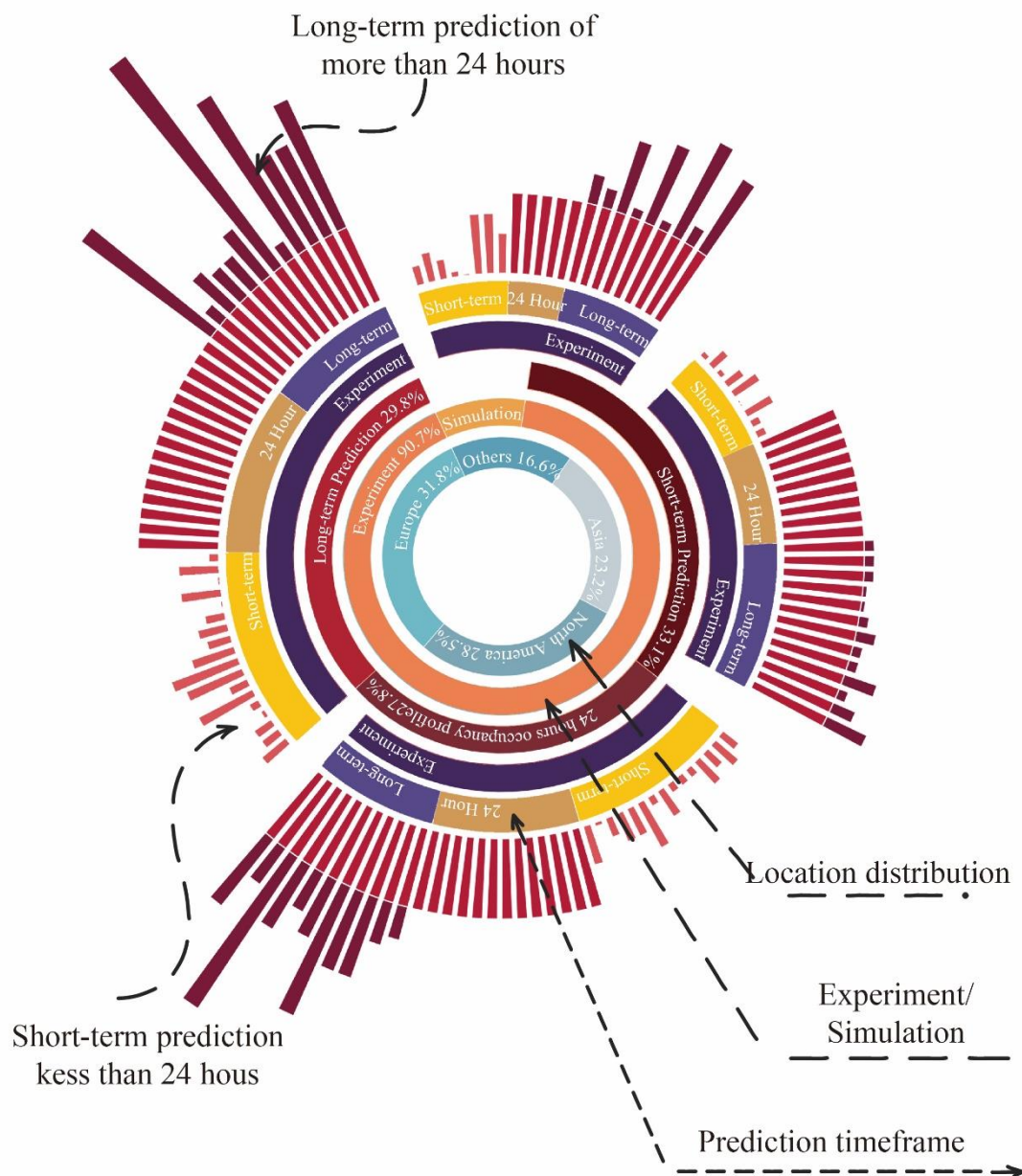


Figure 2-10 The prediction timeframe and experimental method were conducted in different regions based on the reviewed studies.

Accuracy is an important index for evaluating the model's performance and the baseline could be either raw data collected from sensors, or a baseline set before the prediction. However, because of the multiple variables that influence their performance, a straight comparison of the study cases may not be the ideal method. Indeed, models are developed for various places and periods, using data of varying quality, and

supplemented by scripts of varying quality. Even the value used to determine accuracy in different studies differs including mean absolute percentage error, mean percentage error, RMSE, and coefficient of variation of RMSE, making comparison impossible. Table 2-7 shows the algorithms and accuracy index used in some of the reviewed papers, which indicate the different kinds of the mean for accuracy determination used in various models.

Table 2-7. Summary of the algorithm, prediction time and accuracy in some of the reviewed studies.

| Ref. | Year | Prediction time | Algorithm | Accuracy |
|-----------------------------|-------------|------------------------|---|---|
| (Salimi et al., 2019) | 2021 | 30 mins and 5 mins | Inhomogeneous Markov chain | 86% and 68% for lighting and HVAC systems |
| (Chen et al., 2021) | 2021 | 6 months | ANN and fuzzy logic techniques | Reduce the average RMSE by 35% |
| (Chong et al., 2021) | 2021 | 9 months | LSTM | RMSE reduced from 37% to 24% |
| (Apostolo et al., 2021) | 2021 | 24 hours | Multilayer Perceptron ANN | 86.69% accuracy for classification and RMSPE of 0.29 for occupancy counting |
| (Kim et al., 2018) | 2021 | 24 hours | LSTM cells in RNN algorithms | RMSE of 4.48% |
| (Tagliabue et al., 2021) | 2021 | 2 months | Markov model for comfort assessment and ANN for CO ₂ predictions | $R^2 = 0.92$ |
| (Eini and Abdelwahed, 2019) | 2019 | 6 months | ANN | MSE error is 0:003189 |

2.5 Thermal Comfort: Theories, Assessment, and Prediction

Thermal comfort is commonly defined as “that condition of mind which expresses satisfaction with the thermal environment” (ASHRAE, 2019). It is influenced by both

environmental variables—such as air temperature, mean radiant temperature, humidity, and air velocity—and personal factors including metabolic rate and clothing insulation. Ensuring thermal comfort is critical not only for occupant satisfaction and well-being, but also for improving the energy efficiency and responsiveness of building systems.

The understanding of thermal comfort is primarily in two theoretical models: the heat balance theory and the adaptive comfort model. The heat balance approach, often referred to as the rational model, was formalised by (Fanger, 1970b) and central to this model is the Predicted Mean Vote (PMV), which estimates the average thermal sensation vote of a large group of individuals on a seven-point scale ranging from cold to hot. PMV is derived from a steady-state energy balance between the human body and its environment, taking into account six core parameters: air temperature, mean radiant temperature, relative humidity, air speed, clothing insulation, and metabolic rate. While PMV has been widely adopted in standards and simulation tools, it assumes a static indoor environment and homogeneity among occupants, and it does not account for behavioural or psychological adaptation.

In contrast, the adaptive comfort model emerged from extensive field studies and posits that individuals are capable of adapting to their environment through behavioural, physiological, and psychological means (Carlucci et al., 2018). This model suggests that people accept and are comfortable with a wider range of indoor temperatures when they have access to control opportunities, such as operable windows or fans. As a result, thermal expectations in naturally ventilated buildings differ markedly from those in mechanically cooled environments. The adaptive model is particularly relevant in temperate climates and has become increasingly influential in contemporary comfort standards.

Numerous researchers have sought to improve the performance and address the limitations of the PMV model. Adaptive PMV models have been proposed, incorporating an adaptive coefficient to reflect behavioural and psychological

adaptations (Yao et al., 2009). Additionally, two modified PMV models were developed to enhance the predictive accuracy of the original PMV framework (Kim et al., 2015). Individual variances and unique circumstances, such as non-steady-state heat production, were considered in the Metabolic Predicted Mean Vote (MPMV) model (Laouadi, 2022). Another approach focused on the dynamics of body temperature regulation and the environmental impact on thermal comfort across different parts of the human body (Li et al., 2022). However, even under identical fixed thermal conditions, occupants exhibit varying thermal sensations due to factors such as sex, body mass index, time of day, age, and health status (Wu et al., 2023b). These individual characteristics significantly influence subjective thermal comfort. As a result, researchers have increasingly focused on personal comfort models, which predict an individual's thermal comfort response rather than relying on average responses across large populations. Such models are better suited to understanding the unique comfort needs of individual occupants and align with the growing trend toward intelligent and personalised comfort management systems (Talon and Goldstein, 2015).

In recent years, dynamic thermal comfort which account for real-time, individualised responses to changing environmental and physiological conditions has emerged. Unlike static approaches such as the PMV model, or even adaptive models that rely on long-term behavioural adjustment, dynamic thermal comfort models aim to capture the moment-to-moment variability in thermal sensation using continuous data inputs. These models recognise that thermal comfort is not fixed, but influenced by transient factors such as recent activity, emotional state, or microclimatic variations within a room. While some dynamic approaches use physiological signals or wearables, non-invasive methods—particularly those based on thermographic imaging—are increasingly being explored as scalable alternatives. In this context, this thesis contributes to the field by developing a vision-based dynamic thermal comfort prediction model using deep learning. This approach is intended as an initial investigation into the feasibility of real-

time, personalised comfort estimation based on thermal imagery, and lays the groundwork for future integration into adaptive, occupant-aware HVAC control systems.

These theories have been formalised within key international standards that shape building design and operation. ASHRAE Standard 55 incorporates both PMV and adaptive models, allowing practitioners to choose the appropriate method based on the building's ventilation strategy (Olesen and Brager, 2004). European standards such as EN 15251 and EN 16798-1 similarly define acceptable indoor environmental conditions and support both models depending on building type and operational context (Nicol and Wilson, 2010). The implications for public buildings—such as schools, offices, and hospitals—are significant. Strict application of PMV may require more intensive HVAC operation to maintain narrow temperature bands, while the adaptive approach permits broader comfort zones and supports passive or mixed-mode strategies, potentially reducing energy use.

Recent developments in machine learning have provided substitute techniques for personal thermal comfort prediction. Several studies have utilised device-generated data as model inputs to develop high-precision thermal comfort models. These inputs include heart rate (Nkurikiyeyezu et al., 2017), pulse rate (Tamura et al., 2018), oxygen saturation (Xiong et al., 2016), blood pressure (Choi and Yeom, 2017), electroencephalogram (Shan and Yang, 2020), electrocardiogram (Zhang et al., 2017), and skin temperature (Salehi et al., 2020, Cosma and Simha, 2018), all of which exhibit strong correlations with human thermal sensation and comfort. For example, one study proposed a model that measured blood volume pulse, heart rate, electrodermal activity, and hand skin temperature, achieving an impressive accuracy of 95% (Park and Park, 2022). However, wearable sensors required for these approaches are intrusive, as they involve direct physical contact and necessitate that each occupant wears a device, limiting their practicality in real-world scenarios.

Also, recent studies have concentrated on creating non-invasive, vision-based contactless methods for predicting thermal comfort to overcome these limitations. Advances in computer vision technology allow predictions to be made quickly and accurately without directly interfering with users (Zhou et al., 2020). Unlike wearable sensors, cameras—particularly infrared cameras—can capture comprehensive thermal images that include temperature values while preserving occupant privacy (Metwaly et al., 2019b). This approach reduces the need for physical contact and enables more seamless integration into everyday environments.

Table 2-8 Vision-based machine learning research for building thermal comfort in recent years.

| Ref . | Year | Sensor | Model input | Model process | Algorithm | Accuracy |
|---------------------------|------|--|---|---|---|---------------|
| (Ranjana and Scott, 2016) | 2016 | Thermographic camera, air temperature and humidity and radiant temperature | Thermographic data of 7 facial regions and 6 hand regions | Manually obtain temperatures in each region and use as features input into machine learning classifications | Rotation Forests | 94-95% |
| (Burzo et al., 2017) | 2017 | One scientific and one cost-effective thermal camera | Thermal features from faces | Face segmentation, face tracking, and thermal map formation | Decision tree | exceeded 70 % |
| (Ghahramani et al., 2018) | 2018 | Four infrared sensors on an eyeglass frame and room temperature | Ear, nose, front face, and cheekbone temperature | Skin temperature values as input for the machine-learning model | Hidden Markov model | 82.8 % |
| (Li et al., 2018a) | 2018 | Low-cost thermal cameras | Temperature of the forehead, nose, cheeks, ears, lips, and neck | The application of computer vision for human face detection and region of interest extraction; use of statistical methods for the | Haar Cascade algorithm, Kernel smoother, Random | average 85% |

| | | | | | | |
|----------------------------|------|---|--|---|---|-----------------|
| | | | | cleaning and analysis of raw skin temperature data; and implementation of machine learning techniques to create personalised comfort prediction models and examine relevant facial skin temperature features. | Forest classifier | |
| (Jazizadeh and Jung, 2018) | 2018 | RGB video images | The magnified colour values are derived from the pixels representing skin. | Face recognition is used for skin isolation and image magnification. The magnified values of colour are extracted; the thermoregulation state can be identified | Eulerian video magnification algorithm | 89% |
| (Cosma and Simha, 2019b) | 2019 | A colour and a thermographic camera | Generic heat maps | Face detection first identifies the face borders, next computes the generic heat maps and then trains a classifier. | SVM and random forest | 76% |
| (Cosma and Simha, 2019a) | 2019 | a colour camera, a depth sensor, and a thermographic camera | Both skin and clothing temperature | Occupant detection and body parts identification, extract skin and clothing temperatures, the thermal model was trained using the thermal profile of all identified local body parts | SVM, Gaussian process classifier, k-neighbours classifier and random forest classifier | higher than 80% |
| (Salehi et al., 2020) | 2020 | Infrared sensor and a thermostat | Four-point skin temperatures | 14 skin temperature points were measured, and 4 points were selected with a correlation matrix. With chosen points, an optimal method is defined for estimating the thermal sensation | ANN, Decision Trees, Gaussian Process Regression, Fit Regression Ensemble and Group Method of Data Handling methods | 86% |
| (Back et al., 2023) | 2022 | Thermal camera and microclimate data | Pre-processed infrared images | To create an ROI image that only displayed the human body region, all of the photographs were | Deep convolutional neural network | 96% |

| | | | | | | |
|-----------------------|------|---|---|---|---------------|-----------|
| | | | | transformed to greyscale. The pre-processed photos were fed into the prediction model after being scaled to a resolution appropriate for the CNN structure. | | |
| (Jeoung et al., 2023) | 2023 | RGB and thermal camera | Skin temperature of the face | (i) face detection, (ii) skin temperature extraction from ROIs of face components, and (iii) thermal comfort prediction using machine learning models | YOLOv5 | 90.26% |
| (He et al., 2023) | 2023 | Thermal infrared cameras | Cheek, nose, and hand temperatures | Camera imaging, body detection, image registration, data extraction, and modelling | Random Forest | up to 96% |
| (Wu et al., 2023a) | 2023 | Two IR sensors and air temperature distribution | Air, hand, and mean facial temperatures | The data point contains information on skin temperature, environmental parameters, and TSV, which were collected as inputs for the data-driven ML algorithm | Random forest | 0.84 |

Table 2-8 summarises recent research on individual thermal comfort prediction using vision-based methods. A common pattern in these studies involves detecting and tracking occupants within an image, identifying regions of interest (ROIs) (e.g., the face or specific body parts), and applying machine learning algorithms to predict thermal comfort (Burzo et al., 2017). However, a significant limitation of these earlier approaches is their heavy reliance on manual feature extractions specifically, the need for researchers, developers or practitioners to manually select which areas of the image (ROIs) should be analysed for thermal comfort prediction models.

In many studies, traditional machine learning techniques, such as support vector machines (SVM) or decision trees, have been employed to predict thermal comfort. These methods typically require researchers to extract specific temperature data from ROIs, often focusing on localised points of interest, such as the forehead or other visible body parts in infrared images (Li et al., 2018a). Some studies have gone further by

converting infrared images into skin temperature maps before inputting the data into machine-learning models (Jeoung et al., 2023). This preprocessing step ensures more detailed thermal information is available, improving the model's predictive accuracy. However, even with these more comprehensive temperature maps, the process still involves the manual selection of temperature regions, which may limit the system's scalability and its ability to operate effectively in real-time environments.

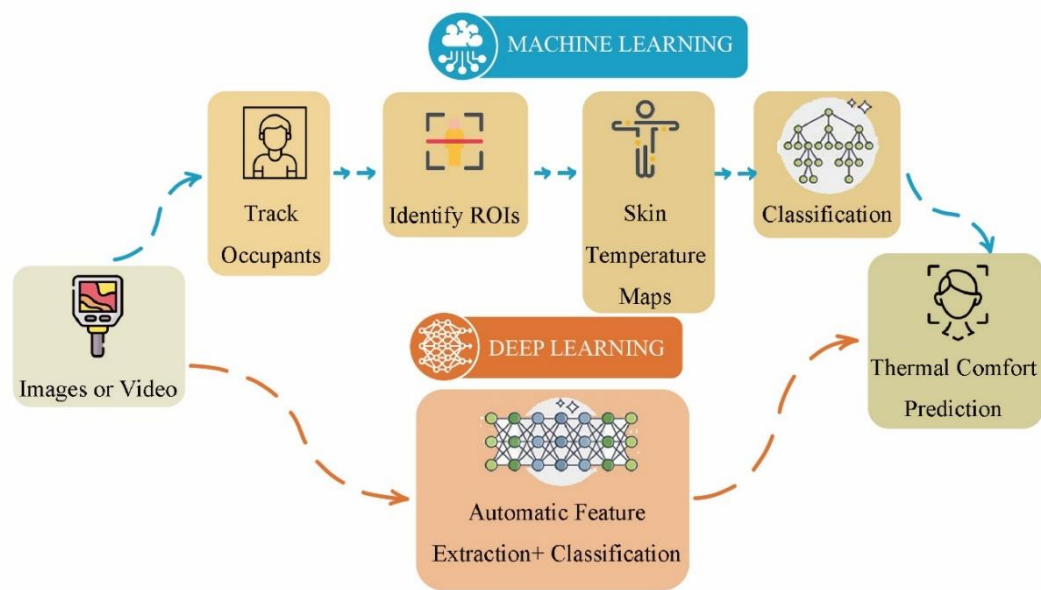


Figure 2-11 Comparison between workflow steps of machine learning and deep learning for vision-based thermal comfort prediction models.

The key distinction between machine learning and deep learning lies in how features and regions of interest (ROIs) are handled (as shown in Figure 2-11). In machine learning-based methods, much of the process depends on manual intervention. For instance, researchers or practitioners often convert infrared images into skin temperature data and manually extract relevant temperature points to input into the model (Salehi et al., 2020). While this approach can provide detailed data, it requires significant human input for feature and ROI selection, making the method less adaptable and automated. Moreover, the accuracy of thermal images can degrade with distance, often necessitating the use of expensive, high-resolution cameras to ensure precision. Additionally, relying

on temperature data from only a few localised points limits accuracy, as it fails to capture the overall thermal distribution across the body.

In contrast, deep learning models, particularly Convolutional Neural Networks (CNNs), have the potential to directly process raw thermal or infrared images without requiring manual conversion into skin temperature maps or the selection of specific ROIs (Alaskar and Saba, 2021). These models are designed to automatically identify key thermal features and regions of interest by analysing the entire thermal image, which could provide a more comprehensive view of the occupant's thermal profile. Such automated feature extraction holds promise for eliminating the need for manual intervention, which can enable real-time thermal comfort prediction and allow the model to continuously learn and improve from new data.

Previous studies have often focused on extracting discrete temperature data points or regions of interest (ROIs), which may limit the model's ability to generalise across diverse conditions and reduce its adaptability to real-world scenarios. Furthermore, most research has not fully explored the potential of deep learning to directly interpret raw thermal images for thermal comfort prediction, leaving a significant gap in the development of fully automated and adaptable systems.

2.6 Research gap

This chapter presents a comprehensive review to explore key advancements in the application of machine learning techniques to building management. Studies generally follow a structured workflow comprising steps such as data collection, algorithm development and application, and validation methods. This workflow was employed by many occupancy prediction methods, enabling dynamic and adaptive building energy management. However, several gaps remain as follows, particularly in the use and evaluation of vision-based methods.

- 1) Vision-based methods, particularly those using thermal cameras, are underexplored compared to traditional sensor-based approaches. Datasets used for occupancy prediction are often limited in diversity and do not include scenarios utilizing vision-based sensors.
- 2) While several camera types (RGB and thermal) are discussed in the literature, their relative performance under varying conditions (e.g., lighting, privacy concerns, cost, accuracy) has not been comprehensively evaluated.
- 3) Privacy concerns remain a barrier to the widespread adoption of cameras in occupancy prediction. Further work is needed to develop and promote privacy-preserving techniques for vision-based methods.
- 4) Few studies systematically compare multiple deep learning algorithms for vision-based occupancy prediction in real-world scenarios.
- 5) With new algorithms emerging frequently, studies often fail to test their suitability for specific building scenarios.
- 6) Current approaches often rely on generalised models for thermal comfort prediction, and occupant-specific thermal comfort often relies on manually defining regions of interest (ROIs) which leaves a gap for more personalized and adaptive solutions.

2.7 Summary

Results of the literature showed that the application of machine learning in building has significantly grown in recent years. The number of studies focusing on occupancy state predictions outnumbered other applications in the early years. The focus of occupancy prediction research is shifting from simply determining whether there are people inside

a room toward more complicated objects such as the occupant's motion, resulting in more accurate building simulation models and better building service operation.

Compared to other commonly used sensors, vision-based sensors do not get enough attention like temperature and CO₂ sensors. This gap is primarily due to privacy concerns associated with cameras, as they capture identifiable visual data. Despite these challenges, vision-based methods, particularly thermal cameras, show significant promise for detailed occupancy detection and thermal comfort modelling. The review also highlights the importance of combining different types of data collection methods and sensors to capture the dynamic variations within buildings and make the necessary adjustments.

Machine learning implementations in different stages of the occupancy prediction workflow were evaluated. One of the most popular algorithms in building occupancy prediction is the neural-network-based algorithm, particularly ANN - LSTM, which was utilised by more than 10 papers after 2018 (Jiang et al., 2021). However, the best method for a specific scenario differs depending on the circumstances. The rapid evolution of machine learning continues to introduce new algorithms with improved capabilities. While these advancements hold great promise, no single algorithm is universally superior.

The review of existing literature reveals advancements in vision-based occupancy detection methods, highlighting the effectiveness of deep learning models such as CNN, YOLO, and Faster R-CNN. Many studies use various types of algorithms, and a comprehensive evaluation of various deep learning models (Dridi et al., 2022) in building environments is yet to be conducted. The lack of a standardised dataset applicable to the building field and the variance in results even when employing the same algorithm underscores a gap in the current body of research (Gursel Dino et al., 2022). Testing different vision-based deep learning models on a consistent dataset within a building

environment would provide valuable comparative insights. Also, CNN and early iterations of YOLO have been extensively utilised in research, yet the latest YOLO algorithms have not been sufficiently evaluated in a realistic and dynamic building environment.

According to the study, investigations on thermal comfort and IAQ prediction using ML are rather limited compared to other domains such as occupancy prediction and energy consumption prediction. According to the study, there is a growing trend of research into occupant comfort and occupancy-centric comfort systems. The concept of thermal comfort is changing from physical index like PMV to occupant's overall comfort, which needs more attention in future works (Xie et al., 2020). Occupants' behaviour, including operating the HVAC system, is driven by their satisfaction with the overall comfort and leads to changes in energy consumption. Therefore, advanced models in the future which maintain comfort and minimize energy consumption will have a promising future. Individual occupant diversity should also be considered, and future models could include exact comfort measures and responses gathered through thermal-based data collection methods such as thermal cameras and thermal comfort rating apps.

3. METHODOLOGY

This chapter presents the methodological framework of the thesis, including the overarching research design, three experiments, sample sizes, data collection procedures, and equipment used. The thesis is positioned within a vision-based, machine learning-driven approach to building performance optimisation, focusing on occupancy detection and thermal comfort prediction.

The research was designed as a three-phase study to systematically investigate the application of vision-based deep learning models for occupancy detection and thermal comfort prediction in buildings. All experiments were integrated into a unified framework aimed at developing practical solutions for occupant-centric building control.

The first phase focused on assessing the performance of various deep learning models for occupancy detection. The second phase compared sensor modalities, evaluating the relative strengths and limitations of RGB and thermal cameras in occupancy prediction. The final phase explored the feasibility of using thermal imaging and deep learning for personalised thermal comfort prediction. The logical progression of these phases allowed for incremental development of model complexity and ensured that the findings from each stage informed subsequent work.

3.1 Research Structure

This thesis is structured as a workflow to systematically address key challenges in vision-based deep learning methods for building management. The workflow begins with an evaluation of different deep learning algorithms for occupancy prediction, followed by a comparison of camera technologies, and concludes with an exploration of thermal comfort modelling as an initial step toward occupant-focused energy management systems. The general workflow for this thesis is shown in Figure 3-1.

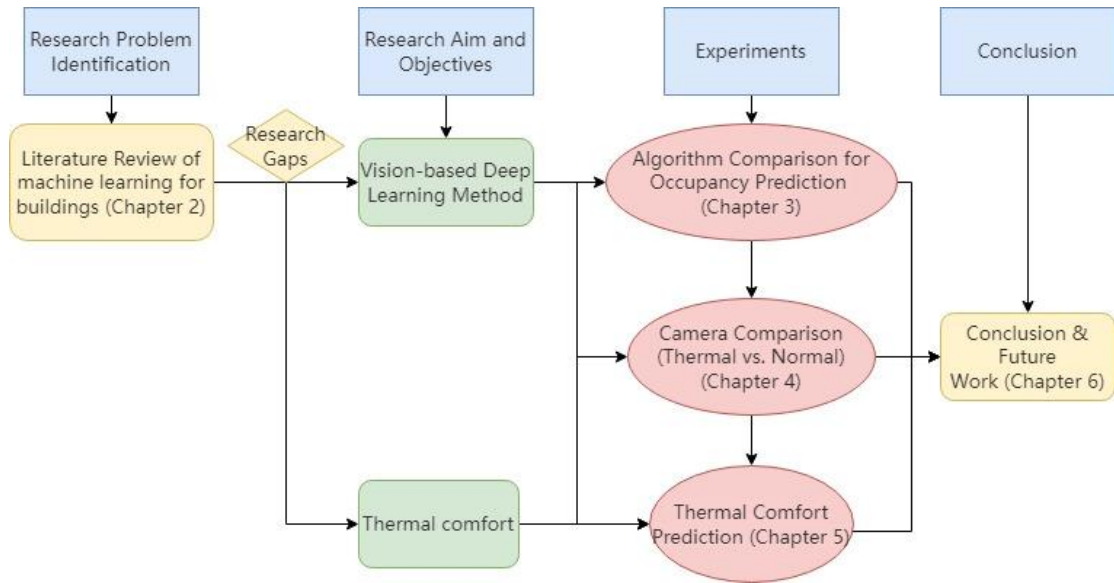


Figure 3-1 The workflow for PhD Methodology in this thesis.

All three experiments were designed to be independent phases of a unified study, each phase building upon the results of the previous one. This integrated approach allowed for development of the vision-based methodology, ensuring that findings from each stage informed and strengthened subsequent analyses. Specifically, the evaluation of occupancy detection algorithms provided a foundation for testing different sensing modalities, which in turn established a basis for investigating occupant thermal comfort through deep learning.

The thesis aims to contribute to the development of vision-based methods for building management by exploring their potential to improve occupancy prediction and thermal comfort modelling. Through systematic experimentation and analysis, the findings provide a deeper understanding of the strengths and limitations of different deep learning algorithms and camera modalities, offering practical recommendations for their application in intelligent building control. Moreover, this study lays the groundwork for future research into integrating real-time, personalised occupant data into adaptive energy management frameworks, further enhancing building efficiency and occupant well-being.

3.2 Experimental Setting

All experiments were conducted at the University of Nottingham campus in various indoor environments, to represent realistic and controlled building conditions.

The research was designed as a three-phase study to progressively address the challenges of vision-based occupancy detection and thermal comfort prediction. The first phase was conducted in a lecture room on the University of Nottingham campus, with areas 96.9 m² and occupancy levels varying from barely occupied to fully occupied scenarios. This phase evaluated different deep learning algorithms for RGB-based occupancy detection, using Logitech C920 RGB cameras positioned at fixed points to record video footage of up to 25 occupants. The environmental conditions, including temperature, humidity, and CO₂ concentration, were continuously monitored using Awair Element sensors to provide context for interpreting the occupancy data.

The second phase built on the first by comparing RGB and thermal cameras to assess their relative performance in occupancy detection tasks. This phase was carried out in similar indoor spaces, including meeting rooms in the Mark Group House, Paton House, and Sustainable Research Building, involving three to eight occupants per test. In addition to the RGB cameras, FLIR ONE Pro thermal cameras were used to capture thermal images alongside RGB video. The duration of each session ranged from 35 to 38 minutes, during which environmental data continued to be recorded. This phase enabled a direct evaluation of sensor modality effects, including detection accuracy, reliability under varying conditions, and privacy implications.

The third phase extended the research to the prediction of individual thermal comfort using thermal imagery and deep learning. Conducted in a temperature-controlled meeting room within the Sustainable Research Building, this phase focused on a single occupant per session. Fourteen participants each took part in an individual session lasting between 90 and 110 minutes. The indoor temperature was dynamically varied

from an initial 12°C to up to 30°C using a Fujitsu split wall-mounted air conditioner operated at low speed to avoid direct airflow effects on comfort perception. Thermal images were collected using a FLIR Lepton 3.5 thermal camera and synchronised with subjective Thermal Sensation Votes (TSVs) reported by the participants at five-minute intervals. Environmental data from the Awair Element sensors were also continuously recorded to provide a complete picture of the indoor climate.

Across all experiments, video data were manually annotated to create bounding boxes for occupancy detection and cropped, greyscaled thermal images for thermal comfort prediction. The details of the settings for the three experiments were listed in Table 3-1.

Table 3-1 Summary of Experimental Settings and Methodology

| | Experiment 1: Occupancy Detection | Experiment 2: RGB vs. Thermal | Experiment 3: Thermal Comfort |
|----------------|---|---|---|
| Purpose | Evaluate deep learning models for RGB-based occupancy detection | Compare RGB and thermal cameras for occupancy detection | Predict individual thermal comfort using thermal images |
| Location | Lecture rooms on campus (96.9 m ²) | Meeting rooms (23.14–53.96 m ²) | Meeting room in Sustainable Research Building |
| Sample Size | Up to 25 occupants | 14 participants, 3–8 occupants per test | 14 participants, individual sessions |
| Equipment | RGB cameras (Logitech C920), environmental sensors | RGB (Logitech C920) and thermal (FLIR ONE Pro) cameras, environmental sensors | Thermal cameras (FLIR Lepton 3.5), environmental sensors, TSV interface |
| Data Collected | RGB video, environmental data | RGB and thermal videos, environmental data | Thermal images, TSVs, environmental data |
| Session Length | 35–60 minutes (depending on scenario) | 35–38 minutes per field test | 90–110 minutes per participant |
| Key Analysis | Model accuracy, precision, recall, mAP | Sensor modality comparison, detection accuracy, privacy aspects | Thermal comfort classification accuracy (intra- and cross-subject) |

3.3 Ethical Considerations

Ethical considerations were central to the design and implementation of all three experiments in this research. All activities involving human participants were conducted in full compliance with the University of Nottingham's ethical guidelines and approved research protocols. Informed consent was obtained from all participants prior to data collection, ensuring that they fully understood the nature and purpose of the study, as well as the data handling procedures in place to protect their privacy.

To maintain participant anonymity, all video and thermal data were anonymised prior to analysis. For occupancy detection tasks, data were handled securely and stored in encrypted formats, with personally identifiable information excluded. In the thermal comfort prediction experiment, thermal imagery was specifically chosen to protect privacy, as it does not contain facial details or other personally identifying visual information. Additionally, in this phase, the split wall-mounted air conditioner in the experimental room was carefully operated at low speeds and positioned to avoid direct airflow towards participants, thus ensuring that no unintended physical discomfort or bias was introduced into the subjective comfort assessments.

By adhering to these consistent ethical practices, this research ensured that participant rights and well-being were safeguarded throughout, while still enabling a comprehensive exploration of vision-based occupancy detection and thermal comfort modelling in real-world building environments.

4. VISION-BASED DEEP LEARNING MODEL COMPARISON FOR OCCUPANCY DETECTION

Some work presented in this Chapter was previously published in the journal [Journal of Building Engineering] as titled *Deep Learning Models for Vision-based Occupancy*

Detection in High Occupancy Buildings by author Wuxia Zhang and co-authors John Kaiser Calautit, Paige Wenbin Tien, Yupeng Wu and Shuangyu Wei. I played a major role in Conceptualization, Methodology, and Writing - the original draft and this study were conceived by all the authors.

4.1 Introduction

This chapter explores the performance of state-of-the-art deep learning models, including SSD, Faster R-CNN, and the latest YOLO series, in a dynamic and realistic building environment, using a low-cost camera setup. By employing single-view camera systems, this research seeks to balance accuracy and cost-effectiveness, providing a practical solution for real-time occupancy detection. The performance of these models will be assessed in terms of speed of detection, computational requirement and capability in complex scenes. A computer vision and deep learning-based approach aimed at detecting and recognising occupants within the building environment was employed. The results from the detection phase are inputs for building energy simulation, which analyses building energy loads and other indices. Figure 4-1 shows the general workflow of the vision-based deep learning model employed in this study.

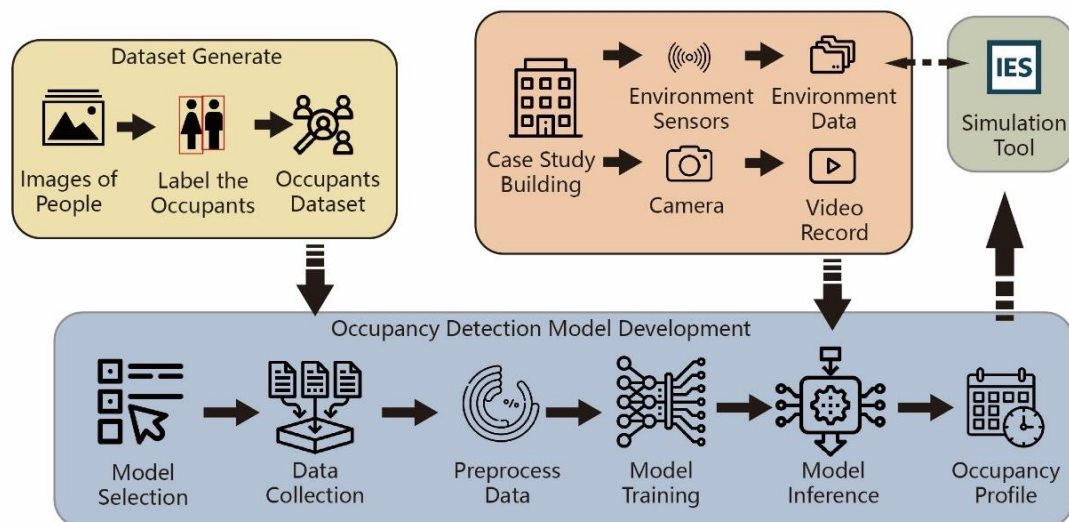


Figure 4-1 The workflow of the proposed vision-based deep learning method for occupancy detection.

4.2 Vision-based occupancy prediction method

4.2.1 Dataset generation

This section outlines the process of creating and preparing the dataset used for training deep learning models. It covers the steps involved in image collection, annotation, and the splitting of the dataset into training, validation, and testing sets. Detailed information is provided on the methodology for generating a standardised dataset, which is essential for reproducibility and understanding the foundation of our model training process.

Deep learning object detection requires a carefully curated image dataset as input. The selection of images should account for various factors to ensure accuracy. For example, variations in indoor lighting, daylight changes, building orientation, and lighting system operations can all affect image recognition (Sun et al., 2020). Therefore, the dataset should include photos from a variety of scenarios, different types of rooms or buildings, varying numbers of people, and multiple angles. These images should be evenly distributed across the test, validation, and training sets to provide comprehensive coverage and improve the model's robustness (Kang et al., 2019).

Since the occupancy dataset in this study is not based on publicly available datasets, the images were manually gathered, annotated, and then randomly divided into three subsets: training, validation, and testing, with a ratio of 88%, 8%, and 4% respectively. A total of 377 images were selected, with 330 for the training set, 31 for the validation set, and 16 for the testing set. The small dataset will allow us to test the capability of the models even with limited data, providing valuable insights into their accuracy and computational efficiency. Our training dataset includes images of humans in diverse environments such as classrooms, offices, and outdoor scenes, to ensure the robustness and generalisation of

the deep learning models. This diversity helps improve the model's performance across different scenarios, making it more versatile and reliable for real-time occupancy detection.

The images for training and validation were sourced from publicly accessible image repositories and manually collected by the research team to ensure diversity and robustness. Notably, the images used for training and validation were not collected from the same room or period as the validation video, which was specifically recorded between 15:15 and 16:21 during a lecture session on December 2nd, 2022. This approach helps to avoid overfitting and ensures that the models can generalise well to different environments.

The collected images were annotated with bounding boxes using the Labelling tool (Tzutalin, 2018), which generated the necessary label files for the training phase. Labelling is an open-source graphical tool used to create bounding boxes, crucial for object detection and image classification tasks. It supports outputs in Pascal VOC XML and YOLO text formats, making it compatible with popular machine-learning frameworks. In this project, 377 images were manually annotated by drawing bounding boxes around each occupant, ensuring high-quality training and evaluation of the deep learning models. The annotation process included loading images, drawing bounding boxes, labelling, and saving annotations in the required format.

Each image was annotated with bounding boxes specifying the exact location of humans, using coordinates for `x_center`, `y_center`, `width`, and `height`. The annotations were saved in `.xml` files for YOLO input and `.txt` files that can be easily converted to TF records for TensorFlow models (Developers, 2022). Pre-processing steps included resizing images to 640x640 pixels and applying auto-orientation to ensure compatibility with the models, preventing memory leaks, poor performance, and imprecise results. Details of the bounding box creation and image preprocessing are listed in Table 4-1.

Table 4-1 Dataset creation and preprocessing steps for occupancy detection

| | Tool | Process | Description |
|-----------------------|-------------------|------------------------|---|
| Dataset creation | Google Images | Image collection | 377 images were collected from Google Images |
| | | Image selection | Images were carefully selected to ensure diversity in postures and arrangements. |
| Bounding box creation | Labellmg | Loading images | Images were loaded into Labellmg for annotation |
| | | Drawing bounding boxes | Bounding boxes were manually drawn around each occupant in the images using the tool's intuitive interface. |
| | | Assigning labels | Each bounding box was labelled to indicate the number of occupants. |
| | | Saving annotations | The created picture annotations were stored as .xml files, which served as YOLO's input, and .txt files that can be easily converted to TF records for TensorFlow models. |
| Image preprocessing | Data augmentation | Resizing | All images were resized to a uniform resolution of 640x640 pixels. |
| | | Auto orient | Ensured images are correctly oriented to prevent memory leaks, poor performance, and imprecise results. |

Figure 4-2 illustrates the variety of images collected and the manual labelling process used to identify the unique regions of interest in each image. This diverse dataset ensures the robustness and generalisation of the deep learning models for occupancy detection across different settings. The number of labels applied to each image was determined by its content. The dataset (https://universe.roboflow.com/wuxia-w5dzu/people_small) has been uploaded and is available on Roboflow, a web-based application for object detection datasets (Wuxia, 2022).

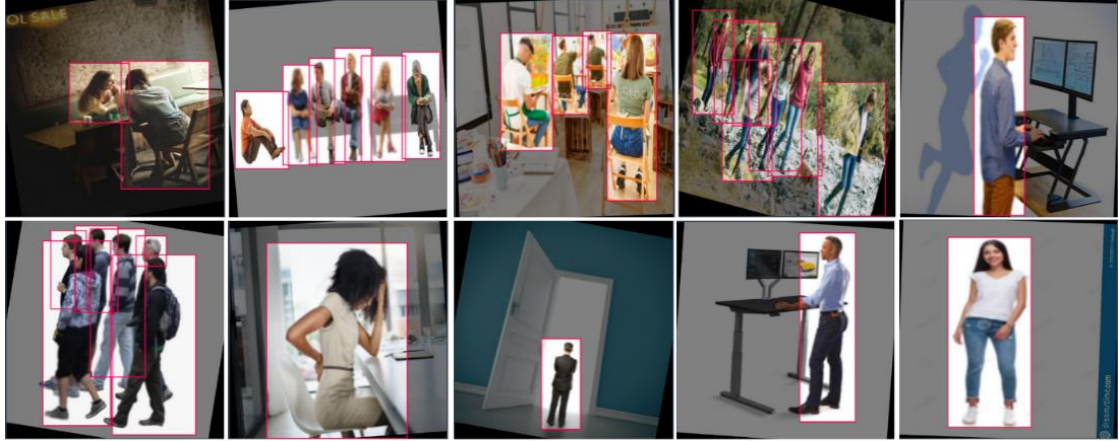


Figure 4-2 Example images from the training dataset, showcasing humans in various environments including classrooms, offices, and outdoor scenes. The diversity of the dataset ensures the generalisation of the deep learning models for occupancy detection in different settings.

4.2.2 Deep learning model training and testing

Given that the experiment is conducted in a real environment characterised by high-density, highly variable, and overlapping occupants, YOLO and SSD have been selected for real-time occupancy detection. SSD MobileNet V2 is recognised for its high speed and reasonable accuracy, attributed to its lightweight nature which facilitates quick inference times, thus making it suitable for real-time detection in high-density spaces (Chiu et al., 2020). YOLOv5 (Choi et al., 2021) has been employed by many researchers while YOLOv7 (Wang et al., 2022) and YOLOv8 (Jocher, 2023) are the latest version and have shown the best performance to date in terms of accuracy and speed. The YOLO series outpaces Faster R-CNN in speed and has shown improved accuracy over earlier YOLO versions, rendering it well-suited for dynamic, high-density spaces. Moreover, it features enhancements geared towards handling smaller objects, which can be advantageous in situations with overlapping occupants (Jocher et al., 2022).

The SSD and Faster R-CNN models were trained using Tensorflow Object Detection on an NVIDIA GeForce GTX 1080 GPU (2560 CUDA cores, 1607 MHz graphics clock, 320GB/s memory bandwidth, 8GB), while the YOLO models were trained with Pytorch in Google Colab (Bisong and Bisong, 2019), which provides free access to NVIDIA T4 Tensor Core GPU (2560 CUDA cores, 1590 MHz graphics clock, 320GB/s memory bandwidth, 16GB).

The decision to use two different GPUs is due to the availability of the GPU on Google Collab. Maintaining a valid and insightful comparison across models, despite using varied hardware, is crucial. Both GPUs have the same CUDA core count and nearly identical clock speeds, which are critical for training speed and computational capacity. They also share a 320 GB/s memory bandwidth, ensuring aligned performance. The primary distinction is the NVIDIA T4's larger memory compared to the GTX 1080, allowing for potentially larger batch sizes or more complex models. Despite this, the comparative analysis remains valid since the core specifications affecting training and inference performance are closely aligned.

All models were trained in the same dataset to ensure consistency. The model loss curves, shown in Appendix A, illustrate the training process and help avoid underfitting or overfitting. Training stopped either when no further improvement was observed or when the loss consistently fell below a certain threshold. SSD and Faster R-CNN required more than 40,000 steps to complete training, while the YOLO models took less than 300 epochs due to their different architectures. The training speed and mAP of the different models is summarised in Table 4-2.

Table 4-2 Comparison of different objection detection models' training performance in this study. The best results for each category are highlighted in bold.

| Model | Year | Platform | Training time (hours) | Epochs | mAP ⁵⁰ (%) | MIT (ms) | GPU |
|---|------|----------------|-----------------------|--------|-----------------------|-------------|-----------------|
| SSD MobileNetV2 (Chiu et al., 2020) | 2020 | Tensorflow1.14 | 8.69 | 48712 | 0.22 | 42 | NVIDIA GTX 1080 |
| Faster R-CNN InceptionV2 (Ren et al., 2015) | 2019 | Tensorflow1.14 | 2.9 | 41901 | 0.83 | 79 | NVIDIA GTX 1080 |
| YOLOv5n (Jocher et al., 2022) | 2020 | Pytorch1.7 | 0.39 | 240 | 0.64 | 28.0 | Tesla T4 |
| YOLOv5x (Jocher et al., 2022) | 2020 | Pytorch1.7 | 0.42 | 240 | 0.63 | 27.6 | Tesla T4 |
| YOLOv7 ^(Wang et al., 2022) | 2022 | Pytorch1.12 | 1.75 | 300 | 0.68 | 57.5 | Tesla T4 |
| YOLOv7w6 (Wang et al., 2022) | 2022 | Pytorch1.12 | 2.03 | 300 | 0.76 | 47.5 | Tesla T4 |
| YOLOv8n (Jocher, 2023) | 2023 | Pytorch2.0 | 0.32 | 88 | 0.82 | 16.4 | Tesla T4 |
| YOLOv8x (Jocher, 2023) | 2023 | Pytorch2.0 | 0.43 | 51 | 0.87 | 292.1 | Tesla T4 |

The Mean Average Precision (mAP) at an Intersection over Union (IOU) threshold of 0.5 was measured for each model using our dataset. Additionally, the mean inference time (MIT) per image in milliseconds (ms) was evaluated to compare the different models. The best results for each metric are highlighted in Table 3. Faster R-CNN outperformed SSD, achieving a mAP of 0.83, ranking as the second-best in terms of mAP among all models. However, since the Faster R-CNN is a two-stage detection model, its detection is not real-time and has a delay in the detection process (Ren et al., 2015). YOLOv8x achieved the highest mAP among all models in 0.43 hours, albeit with a slower inference time. Conversely, YOLOv8n emerged as the fastest model, completing the training process in 0.32 hours with a mAP of 0.82, and featuring the shortest inference time among all models. Given these results, YOLOv8n is selected for a detailed evaluation, which involves validation of the method with all four cameras. The detailed results will be discussed in the following sections.

4.2.3 Case study lecture room, testing and BES modelling

This section describes the setup of the case study room, including the installation of cameras and environmental sensors, the layout, and the conditions under which the experiments were conducted. It provides context for the practical application and testing of the trained models in a real-world environment, demonstrating the feasibility and effectiveness of our approach.

For the implementation of the proposed vision-based deep learning method, lecture room B5 within the Marmont Centre in the University Park Campus, University of Nottingham, UK (Figure 4-3) was selected. The lecture room, located on the first floor of the building, is used by students in the Architecture and Built Environment department for both lectures and tutorial sessions during weekdays. It is also available to students outside of lecture and tutorial periods and on weekends. The room has a capacity of 48 seats and 96.9 m² of floor space, measuring 12.75 m × 7.6 m, with a ceiling height of 2.5 m. Detailed information about the case study room is provided in Table 4-3.

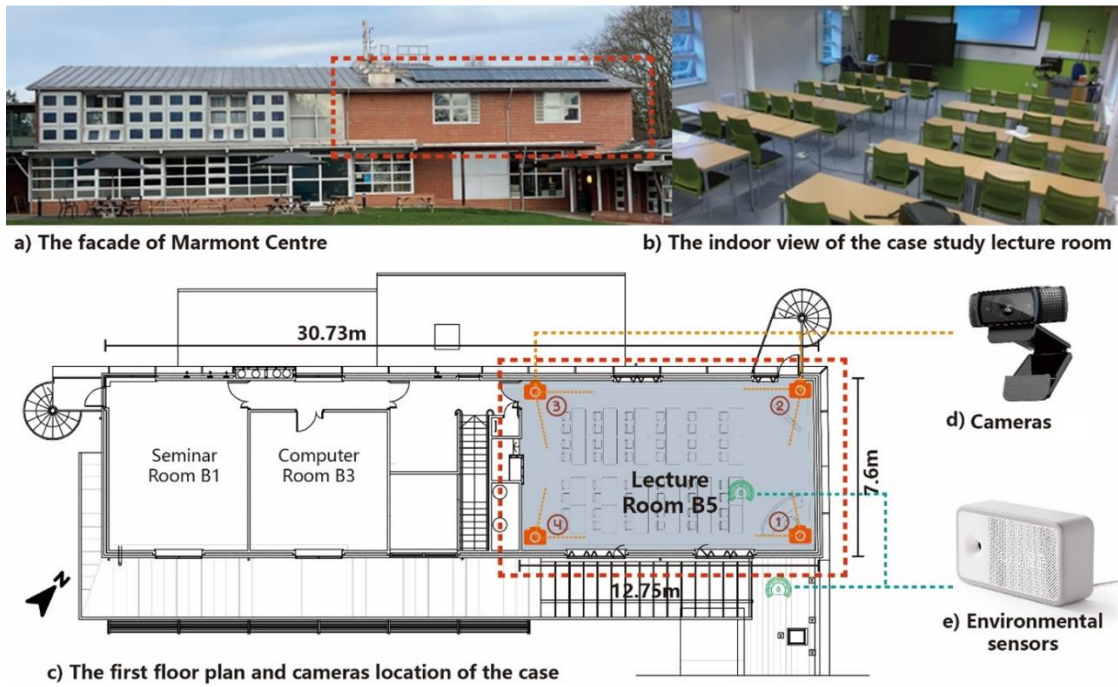


Figure 4-3 (a) The Marmont Centre at the University of Nottingham, UK. (b) The indoor view of the case study lecture room. (c) The floor plan and installed sensors layout of the case study building (d) The picture of the camera in this test. (e) The Awair Element environmental sensors in indoor and outdoor

Table 4-3 Information of the case study lecture room and occupancy profiles.

| | |
|--------------------------------|---|
| Location | Nottingham, UK |
| Room area | 96.9 m ² |
| Room Dimensions | 12.75 m × 7.6 m × 2.5 m |
| Seats | 48 |
| Heating setpoint | 21°C |
| Occupancy schedule (base case) | 08:00 – 18:00 |
| Ground Truth occupancy profile | The observed occupants' number |
| Occupancy detection profile | Profile generated from vision-based occupancy detection |

Four Logitech C920 cameras were installed, one in each corner of the room, capable of recording full-HD 1080p video at 30 frames per second (fps) with a 78-degree field of

view. Additionally, two Awair Element environmental sensors were placed both outside and inside the room to monitor temperature, relative humidity, carbon dioxide levels, volatile organic compounds (VOCs), and fine particulate matter (PM_{2.5}). All environmental sensors were set to record data continuously on December 2nd, 2022. The layout and location of the sensors are shown in Figure 4-3, and detailed information about the sensors used is listed in Table 4-4.

Table 4-4 Environmental sensors and cameras used in the case study experiment.

| Measurement parameters | Sensor | Range | Resolution | Accuracy | Number and location |
|-------------------------------|-------------------------------------|-------------------------|--|--------------|-------------------------------|
| Air temperature | Awair Element environmental sensors | 0-90°C | 0.015°C | ±0.2°C | 2, one inside and one outside |
| Relative humidity | | 0%-100% | 0.01% | ±2% | |
| CO ₂ concentration | | 400 to 5000ppm | 1ppm | 75ppm or 10% | |
| Camera | Logitech C920 | 78-degree field of view | Full-HD 1080p video 30 frames per second | - | 4 in each corner |

The experiment was conducted during a single lecture session, capturing various occupancy conditions: the initial period when participants were entering the lecture room, resulting in barely occupied conditions; during the lecture when the room was occupied; and the period when participants were leaving the lecture room, resulting in barely occupied conditions again. By evaluating the models under these different occupancy levels, we were able to assess their performance in adapting to changing conditions within a single session. During the test, the lecture room was mainly occupied from 15:30 to 16:10, with no other activities scheduled for the day. To evaluate the performance of the trained occupancy detection models in a real-world setting, video recording in the room was carried out from 15:15 to 16:20, focusing particularly on the lecture period with a maximum of 25 attendees. Additional environmental factors, such as relative humidity, temperature, and CO₂ levels, were monitored throughout the day

to capture the conditions during both occupied and unoccupied periods. All participants were students at the University of Nottingham and were informed about the experiment; they consented to the usage of the footage for this study. The detailed experiment workflow is shown in Figure 4-4.

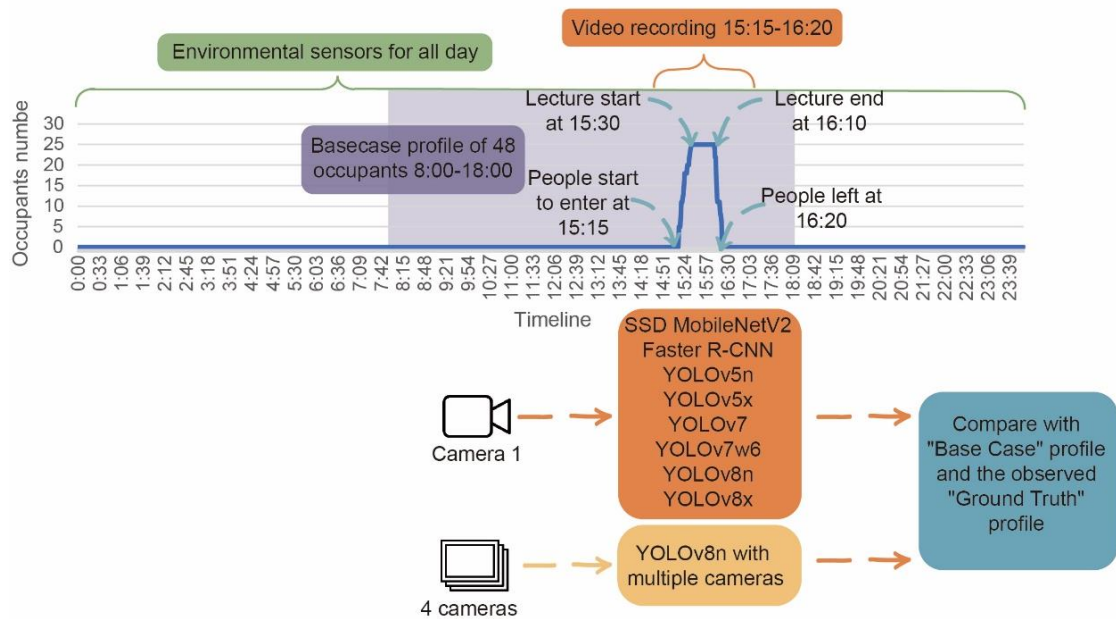


Figure 4-4 The workflow for the different cameras employed in the case study experiment on Dec 2nd, 2022.

Given that video footage from a single viewpoint may suffer from overlapping views and signal synchronisation issues (Hsu et al., 2020), this study initially tests all deep learning models using the recorded video from Camera 1. Subsequent experiments will employ the algorithm demonstrating the best performance, utilising videos from the remaining cameras to assess the influence of different viewing angles. Detailed results from these experiments will be discussed in the following section.

The performance of vision-based models can be significantly affected by various factors, particularly when deployed in complex indoor environments. Although the video footage represents a full lecture scenario, it may not capture all possible occupancy patterns.

Factors such as changing lighting conditions, shadows, and the presence of other objects may lead to inaccurate results. Additionally, occupants situated in corners might be easily missed due to the camera's resolution. While the use of low-cost cameras might constrain accuracy in certain scenarios, it enables cost-effective implementation, making it feasible to deploy these systems without heavily investing in new infrastructure. This approach allows for more widespread adoption of occupancy detection technologies within budget constraints.

Table 4-5 IES Modelling construction details including U-values ($\text{W/m}^2\text{K}$) and thickness.

| | Wall | Roof | Ground floor | Window | Door |
|------------------------------------|------|------|--------------|--------|------|
| U-value ($\text{W/m}^2\text{K}$) | 0.33 | 0.22 | 0.32 | 2.95 | 2.30 |
| Thickness (mm) | 300 | 290 | 230 | 20 | 40 |

IES VE, a Building Energy Simulation (BES) tool, was used to model the building (Solutions, 2020) to evaluate the potential of the proposed approach and to assess its impact on building energy loads and CO_2 concentration predictions. The building is equipped with a central heating system and has operable windows for natural ventilation. During hours when the building is occupied (08:00 – 18:00), the heating system was set to maintain an indoor temperature of 21°C (CIBSE, 2021). In this case study, operational hours of 08:00 – 18:00 on working days (CIBSE, 2008) were assumed for the base case occupancy profile (fixed schedule). For the simulation, a weather data file from Nottingham, UK was used. The respective U-values for the wall, roof, ground, window, and door are detailed in Table 4-5.

4.3 Experiment results and discussion

The following section presents the results, analysis, and assessment of the model detection performance and the impact of the suggested approach on building energy

predictions. The trained models are applied in the case study lecture room to evaluate their performance in a real-world environment.

4.3.1 Comparison of state-of-the-art deep learning models in occupancy detection

Most models except SSD demonstrated the capability to identify, classify, and locate occupants within the lecture room. However, due to varying frame rates and inference speeds among different models, synchronisation across all videos was not achieved. For instance, Faster R-CNN does not facilitate real-time object detection, thereby exhibiting delays, whereas the YOLO series operates in real time without delays.

Video 4-1 illustrates the first 15 seconds of the inference videos from all models, showcasing their performance as individuals began entering the room. As seen in Figure 4-5, while the number of occupants was limited, most models captured all individuals present in the video, except for SSD, which only detected one occupant near the camera—these findings align with its low mAP score as shown in Table 4-2. Occasionally, the deep learning models generated false-positive results, likely due to the presence of objects with patterns resembling the target [56]. For example, YOLOv5n and YOLOv5x falsely identified two objects on the left wall, YOLOv7 positioned a bounding box on the right wall, and YOLOv7w6 misclassified the overhead projector as occupants.

These false positives typically had lower confidence scores, often falling below 0.3, indicating a lack of certainty in those detections and suggesting a potential avenue to enhance model accuracy.

Figure 4-6 shows the scene at 15:45 when all students have settled into their seats and exhibit minimal movement, providing a clearer perspective on model performance in a relatively static scenario. Video 4-2 compares the detection performance from 15:45:00 to 15:45:15. The SSD model's performance remains suboptimal, detecting only one

occupant near the camera. While other models successfully capture the occupants, "flickering" bounding boxes are observed across all models. This flickering, which occurs even in the absence of movement, is a common challenge in object detection algorithms applied to videos. It often arises due to changes in object position, lighting variations, and the underlying algorithmic architecture (Azulay and Weiss, 2018).

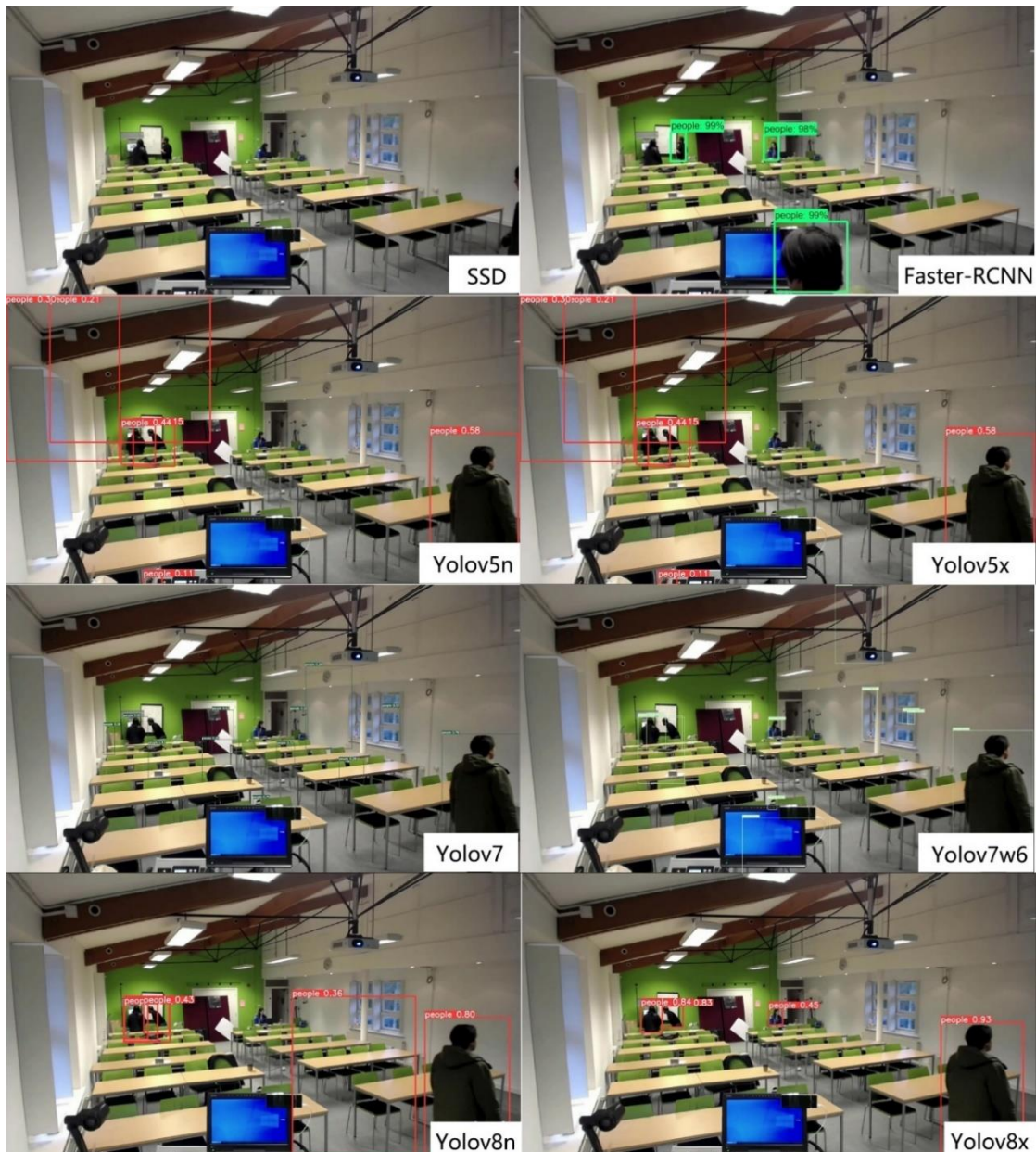
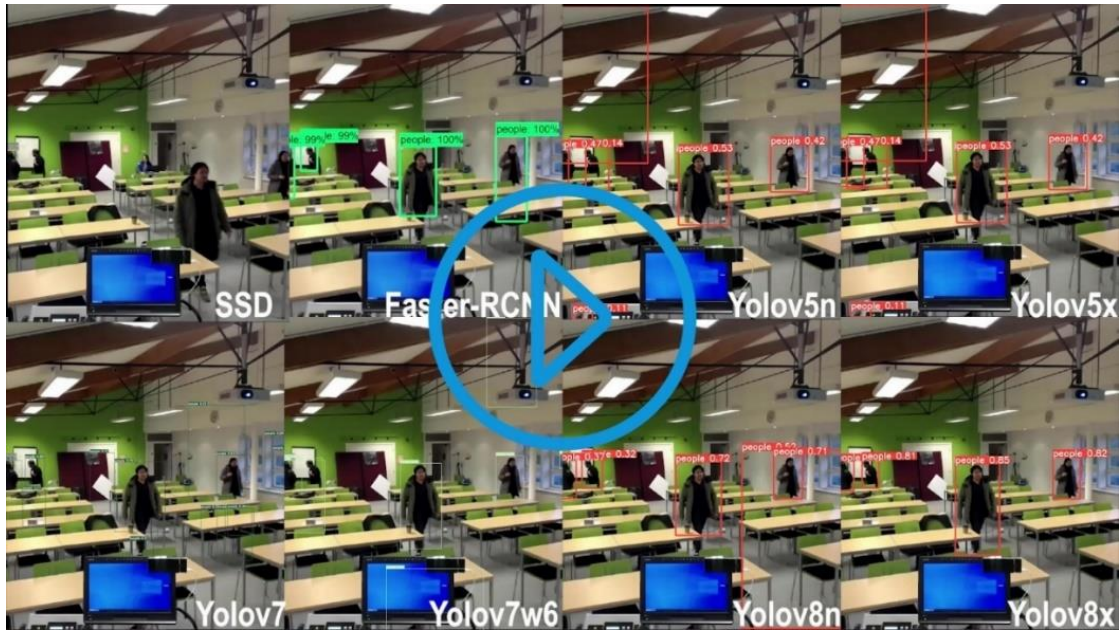


Figure 4-5 The frame at 15:15:10 compares the deep learning models' detection of the participants entering the room.



Video 4-1 The inference videos from 15:15:00 to 15:15:15 compare the detection performance of the models when people were entering the room. *The playable video is available at <https://ars.els-cdn.com/content/image/1-s2.0-S2352710224029231-mmcl.mp4>*

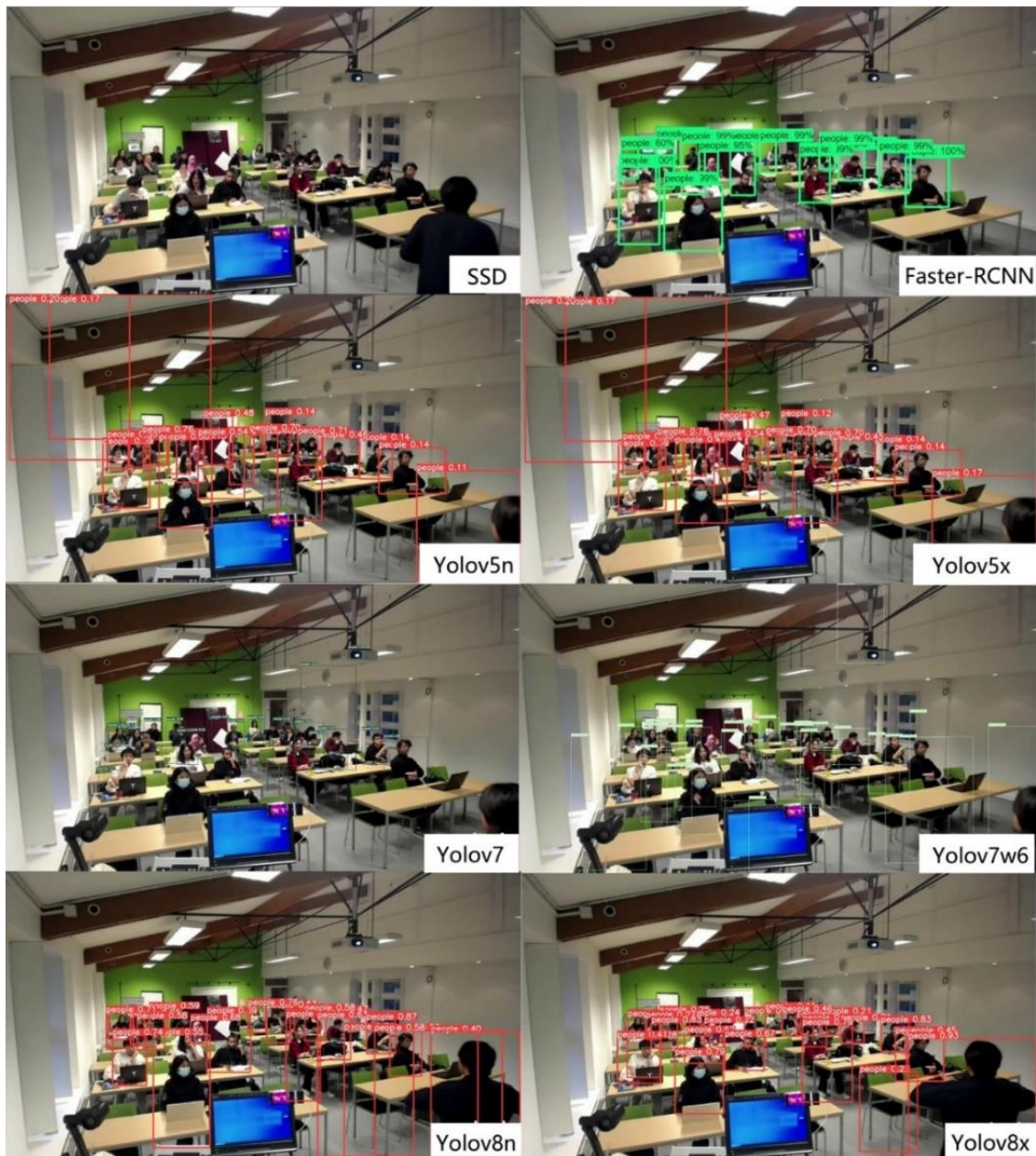
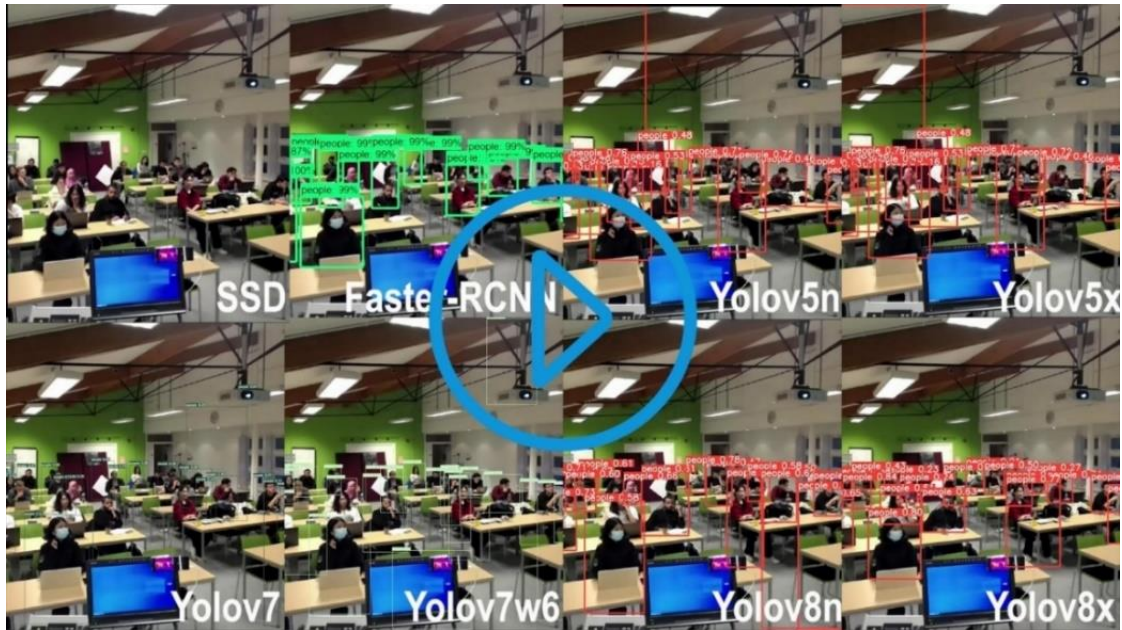


Figure 4-6 The frame at 15:45:00 compares the deep learning model detection of the participants in the middle of the lecture.



Video 4-2 The inference videos from 15:45:00 to 15:45:15 compare the detection performance of deep learning models when the participants were mostly sitting. *The playable video is available at <https://ars.els-cdn.com/content/image/1-s2.0-S2352710224029231-mmc2.mp4>*

In this scenario, YOLOv5n and YOLOv5x continue to exhibit two false positives on the left side, while YOLOv7 and YOLOv7w6 misidentify the screen and overhead projector as occupants. In contrast, Faster R-CNN, YOLOv8n, and YOLOv8x demonstrate a more accurate capture of almost all occupants, although they still produce a few false positives.

Video 4-3 and Figure 4-7 show the scene as students exit the lecture room, with most occupants congregating near the door and moving out of the camera's view. The SSD model continues to fail to detect any occupants, a consistent issue observed in previous scenarios. Faster R-CNN occasionally misses an occupant near the door in certain frames, possibly due to low resolution. YOLOv5n and YOLOv5x mistakenly label two boxes in the left corner—a recurring error—while YOLOv7 generates several false positives on the left seats. YOLOv7w6 and YOLOv8x successfully capture all occupants, though one false positive occurs for the monitor, likely due to reflection.

After discussing occupancy detection in different lecture room scenarios, a detailed analysis of these models' real-world applicability and precision will be conducted. This leads us to a deeper examination using standard metrics in the following section.

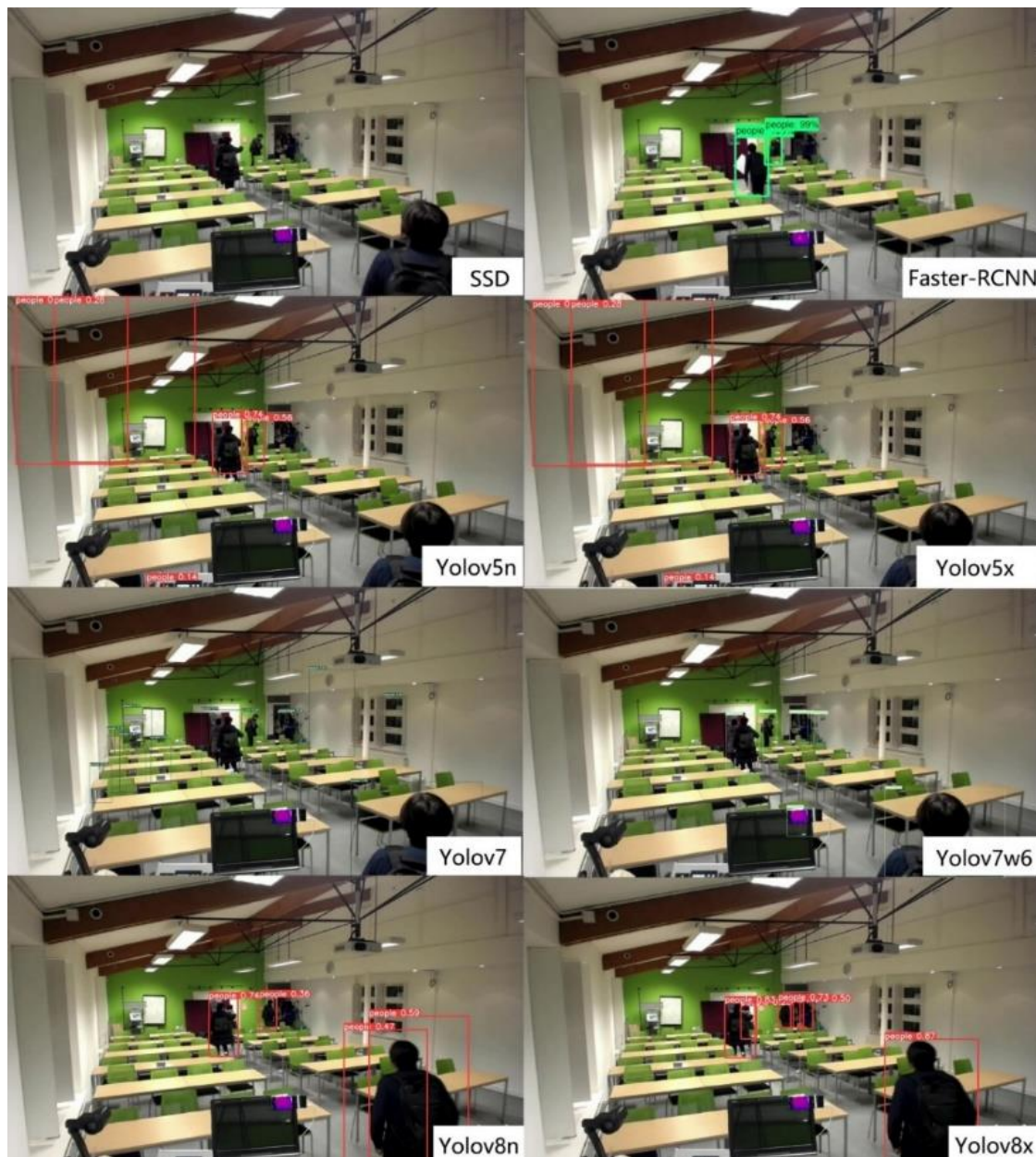
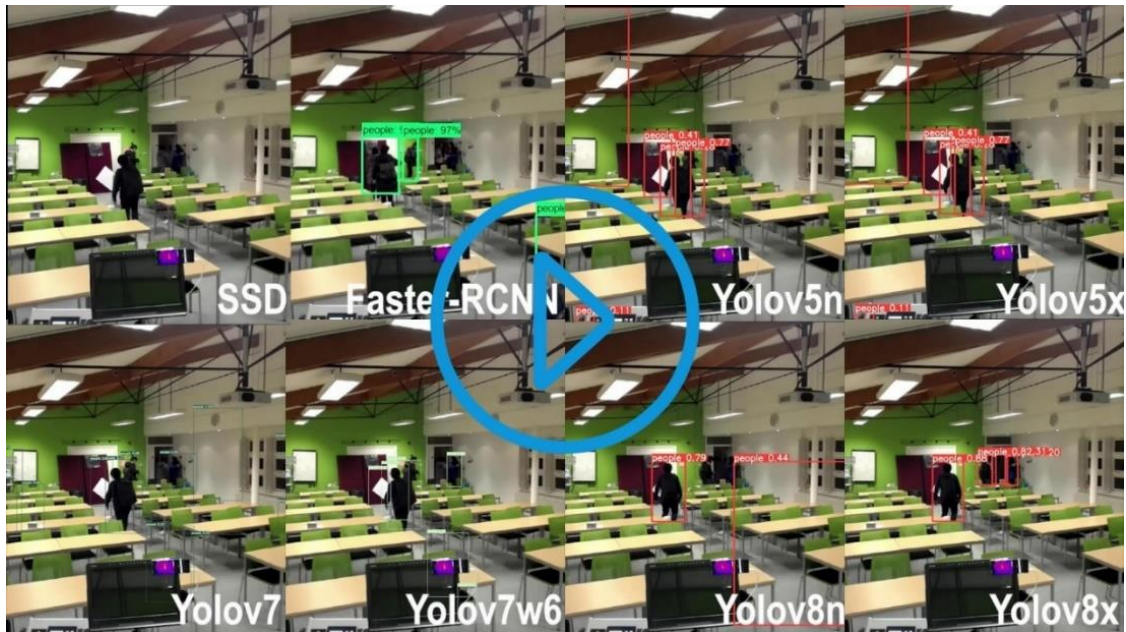


Figure 4-7 The frame at 16:09:40 compares the deep learning models' detection of participants leaving the room.



Video 4-3 The inference videos from 16:09:45 to 16:10:00 compare the detection performance of deep learning models when participants are leaving the room. *The playable video is available at <https://ars.els-cdn.com/content/image/1-s2.0-S2352710224029231-mmc3.mp4>*

4.3.2 Evaluation of the model performance in the case study building

Building on the initial findings from the previous section, this section conducts a detailed evaluation of the models. We use common metrics such as Accuracy, Recall, Precision, and the F1 Score for a comprehensive analysis of each model's performance. Understanding and applying these metrics is crucial, as they provide a quantitative measure of the model's ability to accurately identify and classify occupants within the building environment. Additionally, we explore the impact of camera positioning and different angles on detection accuracy to gain a more comprehensive understanding of the models' practical performance.

The accuracy metric (Sokolova et al., 2006) provides the proportion of correctly classified samples to all samples, offering a broad understanding of the model's performance. The formula for Accuracy is expressed in Eq. (1):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3-1)$$

Where *TP* (true positive) represents the number of correctly predicted occupants in the video, *FP* (false positive) represents the number of predictions where other objects are regarded as occupants, *FN* (false negative) represents the number of undetected occupants, and *TN* (true negative) represents the number of images without occupants where no prediction is performed. Figure 4-8 illustrates examples of True Positive (TP), False Positive (FP), and False Negative (FN) occurrences in a frame captured from the YOLOv7 model inference video at 15:41:40.

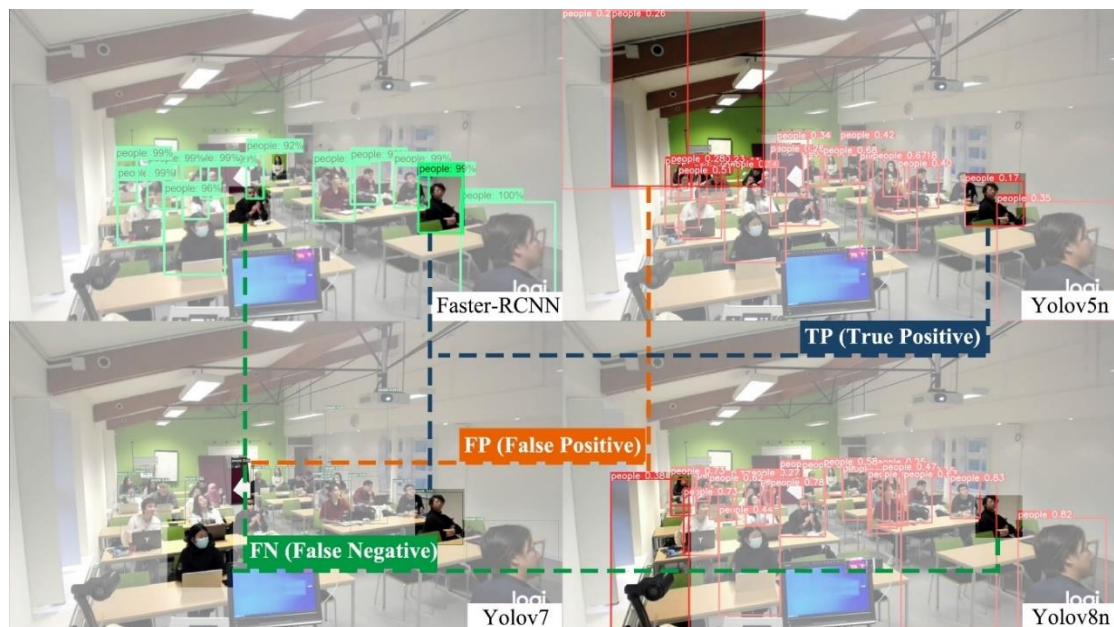


Figure 4-8 Examples of TP, FP and FN in a frame taken from the result of Faster R-CNN, YOLOv5n, YOLOv7 and YOLOv8n at 15:35:00.

The recall is crucial because it displays the proportion of true-positive predictions to all occupants found, which is particularly relevant in scenarios where missing an occupant detection is undesirable. The formula for Recall is expressed in Eq. (2):

$$Recall = \frac{TP}{TP + FN} \quad (3 - 2)$$

The ratio of true positive predictions to all positive predictions made is calculated by precision, while the F1 Score provides a balanced measure between precision and recall, which is particularly useful when we want to understand the model's balance between these two metrics. The formulas for Precision and F1 Score are expressed in Eq. (3) and Eq. (4):

$$Precision = \frac{TP}{TP + FP} \quad (3 - 3)$$

$$F1 \text{ score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3 - 4)$$

Table 4-6 compares the performance of eight deep learning models (SSD, Faster R-CNN, YOLOv5n, YOLOv5x, YOLOv7, YOLOv7w6, YOLOv8n, YOLOv8x) across different metrics (Accuracy, Recall, Precision, and F1 score) during three phases: as participants enter the room, during the lecture, and as they exit the room.

During the first 15 minutes as participants enter the room, YOLOv8x shows the highest accuracy at 0.89, significantly outperforming other models. It also achieves the highest recall at 0.94, indicating its effectiveness in identifying true positive instances. Both Faster R-CNN and YOLOv8x maintain high precision (1.00 and 0.94, respectively), meaning they have the lowest false positive rates. YOLOv8x again stands out with the highest F1 score of 0.94, demonstrating its balanced performance in both precision and recall. In contrast, SSD shows the lowest performance in this phase, with an accuracy of 0.05 and an F1 score of 0.10.

During the lecture, which lasts from 15:30 to 16:10, YOLOv8x and YOLOv8n continue to lead in performance. YOLOv8x achieves an accuracy of 0.75 and a recall of 0.80, while YOLOv8n follows closely with an accuracy of 0.73 and a recall of 0.78. Both models also maintain high precision, with YOLOv8x at 0.93 and YOLOv8n at 0.92. The F1 scores for YOLOv8x and YOLOv8n are 0.86 and 0.84, respectively, indicating their robust performance during the lecture phase. Faster R-CNN, while maintaining perfect precision (1.00), falls behind in recall (0.57) and overall accuracy (0.57).

In the last 10 minutes as participants leave the room, YOLOv8n achieves the highest accuracy at 0.78 and the highest recall at 0.82. It also maintains a high precision of 0.93 and an F1 score of 0.88, making it the best performer during this phase. YOLOv8x, while slightly behind YOLOv8n, still performs well with an accuracy of 0.63, recall of 0.71, precision of 0.86, and F1 score of 0.77. In this phase, SSD performs the poorest with accuracy, recall, precision, and F1 score all at 0.00.

When considering the overall performance across all phases, YOLOv8x emerges as the top performer with an accuracy of 0.77, recall of 0.82, precision of 0.93, and an F1 score of 0.87. YOLOv8n also shows strong overall performance with an accuracy of 0.72, recall of 0.78, precision of 0.90, and an F1 score of 0.84. Faster R-CNN excels in precision (1.00) but lags in recall (0.61) and overall accuracy (0.61). SSD, on the other hand, consistently shows the lowest performance metrics.

YOLOv8x consistently outperforms other models across all phases in terms of accuracy, recall, precision, and F1 score. However, YOLOv8n also demonstrates a balanced and strong performance and is particularly noteworthy for its good speed, making it a highly suitable model for real-time applications where both performance and efficiency are crucial. The superior capabilities of these advanced YOLO models in real-time occupancy detection systems provide valuable insights for their application in building

and energy management, with YOLOv8n being a promising candidate for future use due to its excellent balance of accuracy and speed.

Table 4-6 Comparison of the performance of the deep learning models during three distinct phases: as participants enter the room, during the lecture, and as they exit the room

| Model | | SSD | Faster R-CNN | YOL Ov5n | YOL Ov5x | YOL Ov7 | YOLO v7w6 | YOL Ov8n | YOL Ov8x |
|-----------------------|-----------|------|--------------|----------|----------|---------|-----------|----------|-------------|
| Enter (First 15 mins) | Accuracy | 0.05 | 0.76 | 0.51 | 0.47 | 0.49 | 0.50 | 0.66 | 0.89 |
| | Recall | 0.05 | 0.76 | 0.66 | 0.62 | 0.68 | 0.74 | 0.77 | 0.94 |
| | Precision | 1.00 | 1.00 | 0.69 | 0.66 | 0.64 | 0.61 | 0.82 | 0.94 |
| | F1 score | 0.10 | 0.86 | 0.67 | 0.64 | 0.66 | 0.67 | 0.80 | 0.94 |
| Lecture (15:30-16:10) | Accuracy | 0.03 | 0.57 | 0.57 | 0.46 | 0.57 | 0.60 | 0.73 | 0.75 |
| | Recall | 0.03 | 0.57 | 0.65 | 0.54 | 0.64 | 0.73 | 0.78 | 0.80 |
| | Precision | 1.00 | 1.00 | 0.81 | 0.75 | 0.84 | 0.77 | 0.92 | 0.93 |
| | F1 score | 0.05 | 0.72 | 0.72 | 0.63 | 0.73 | 0.75 | 0.84 | 0.86 |
| Leave (Last 10mins) | Accuracy | 0.00 | 0.59 | 0.50 | 0.41 | 0.56 | 0.43 | 0.78 | 0.63 |
| | Recall | 0.00 | 0.59 | 0.65 | 0.53 | 0.82 | 0.67 | 0.82 | 0.71 |
| | Precision | 0.00 | 1.00 | 0.69 | 0.64 | 0.64 | 0.56 | 0.93 | 0.86 |
| | F1 score | 0.00 | 0.74 | 0.67 | 0.58 | 0.72 | 0.61 | 0.88 | 0.77 |
| Overall | Accuracy | 0.03 | 0.61 | 0.55 | 0.46 | 0.55 | 0.56 | 0.72 | 0.77 |
| | Recall | 0.03 | 0.61 | 0.66 | 0.56 | 0.66 | 0.73 | 0.78 | 0.82 |
| | Precision | 1.00 | 1.00 | 0.77 | 0.72 | 0.77 | 0.71 | 0.90 | 0.93 |
| | F1 score | 0.06 | 0.75 | 0.71 | 0.63 | 0.71 | 0.72 | 0.84 | 0.87 |

Figure 4-9 compares the accuracy of all models with their training and inference times, where a model with better accuracy occupies a larger circular area. The SSD model's performance is notably poor in most frames, even with the longest training time. YOLOv8x, with an acceptable training time, demonstrated superior performance, achieving the highest overall accuracy of 0.77 and an F1 score of 0.87, although its inference time is lengthy. It can detect most occupants but occasionally misses some at the far end. It is worth noting that Faster R-CNN achieved high precision and confidence scores, as shown in Figure 4-9, despite requiring more training time. YOLOv5n,

YOLOv5x, and YOLOv8n all exhibited good performance, recognizing most occupants, although they sometimes mistakenly identified other objects as occupants initially. YOLOv8n has better accuracy, which is why it was selected for the next stage of testing with cameras from various angles. YOLOv7 and YOLOv7w6 provided the same accuracy and F1 score, although their training times and mAP were quite different. These two models occasionally mistook the screen on the table and the overhead projector for occupants.

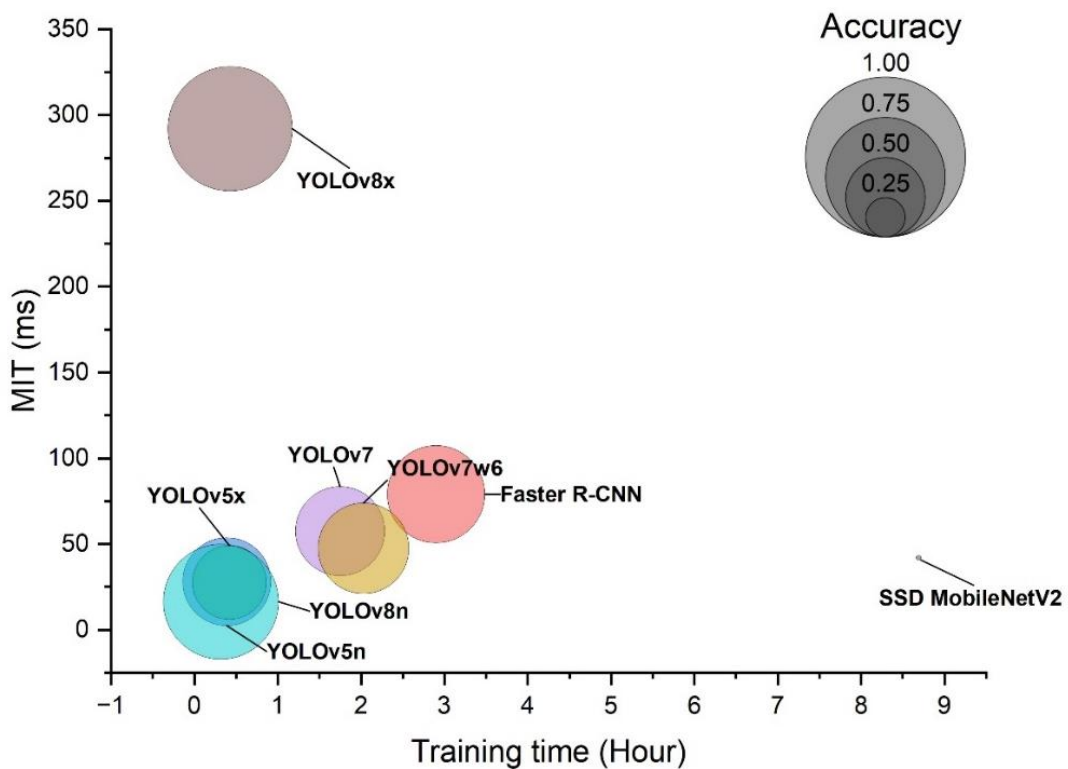


Figure 4-9 Performance of deep learning models comparing training time, inference time and accuracy (higher accuracy model occupied bigger circular area)

Figure 4-10 presents occupancy profiles predicted by deep learning models compared to the actual number of occupants (ground truth). These profiles will be used as input for subsequent energy simulations. The results from the deep learning models represent the number of detected occupants, including both true positive and false positive detections. Consequently, there may be instances where the results approximate the

ground truth but exhibit discrepancies due to missed occupants or false detections. Further work is needed to improve the accuracy, stability, and dependability of the detection models. Based on a comparative analysis of the eight models used in the experiment, YOLOv8n exhibits the least variation in prediction error, yielding the most accurate results when compared to the ground truth.

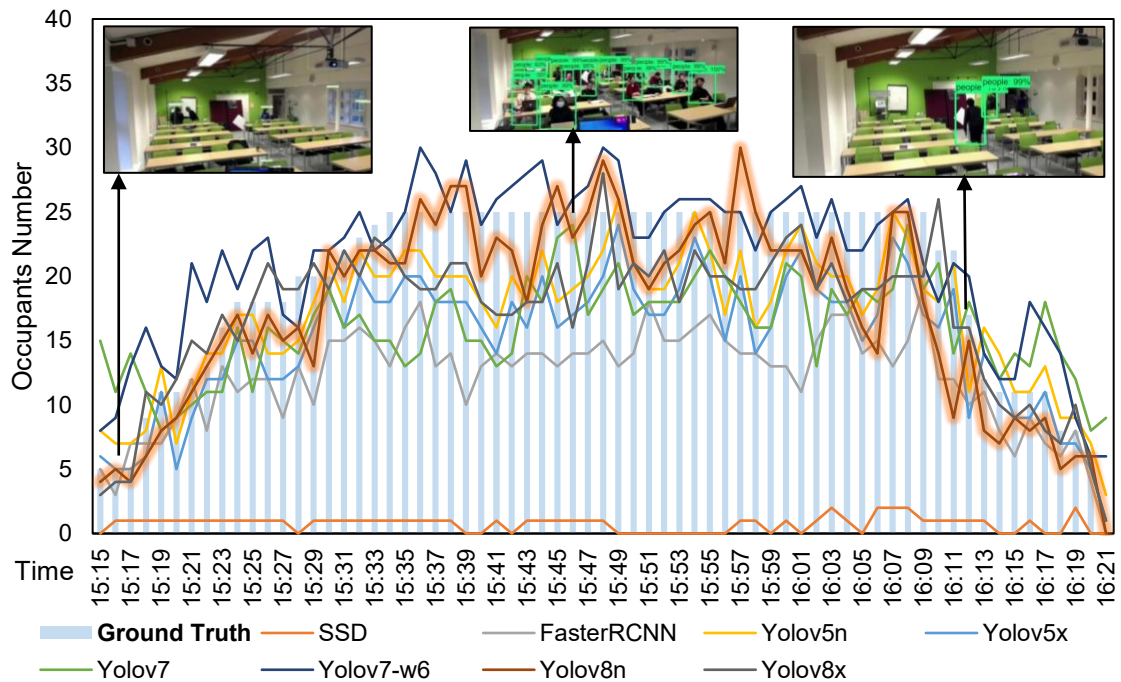


Figure 4-10 The occupancy profiles predicted by the deep learning models, compared to the ground truth.

Moreover, four cameras were strategically positioned at each corner to assess the impact on detection performance, addressing the inevitable occlusions inherent in a singular view. Figure 4-11 shows a frame from the inference videos captured by different cameras at 15:45, a time when most students were seated with minimal movement. Video 4-4 compares the detection from the four different cameras from 15:45:00 to 15:46:00. The YOLOv8n model was selected for this experiment due to its superior overall performance demonstrated earlier. While some occupants were missed by one camera, they were discernibly captured by others. For instance, camera 2 overlooked

several individuals in the right corner, yet they were clearly captured by cameras 1 and 3. This highlights the suboptimal performance of camera 2, which consistently missed occupants in the right corner.

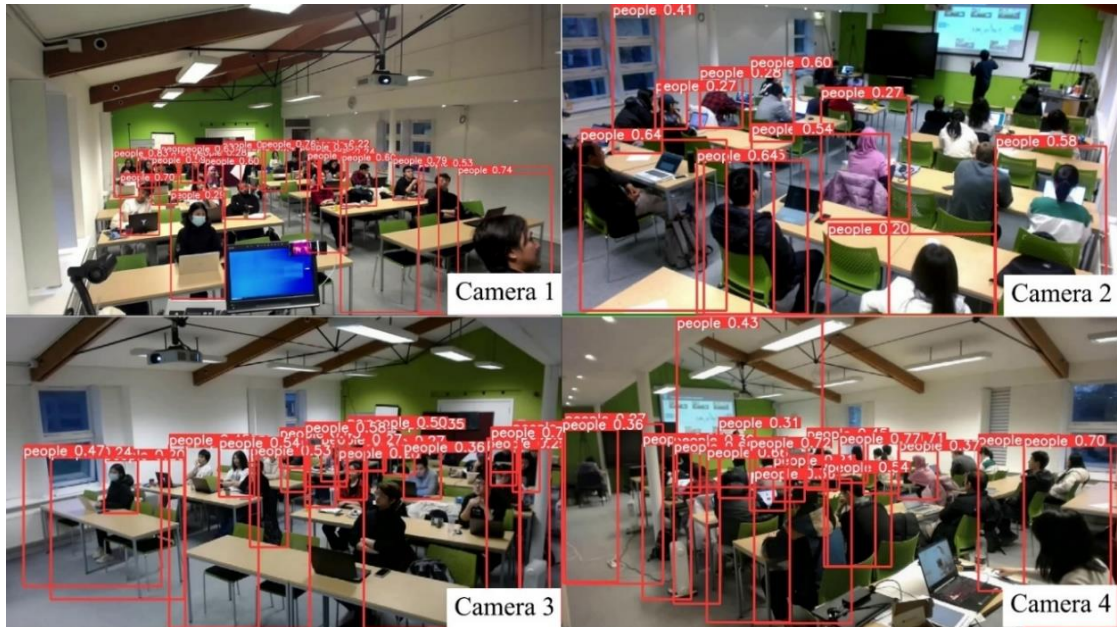
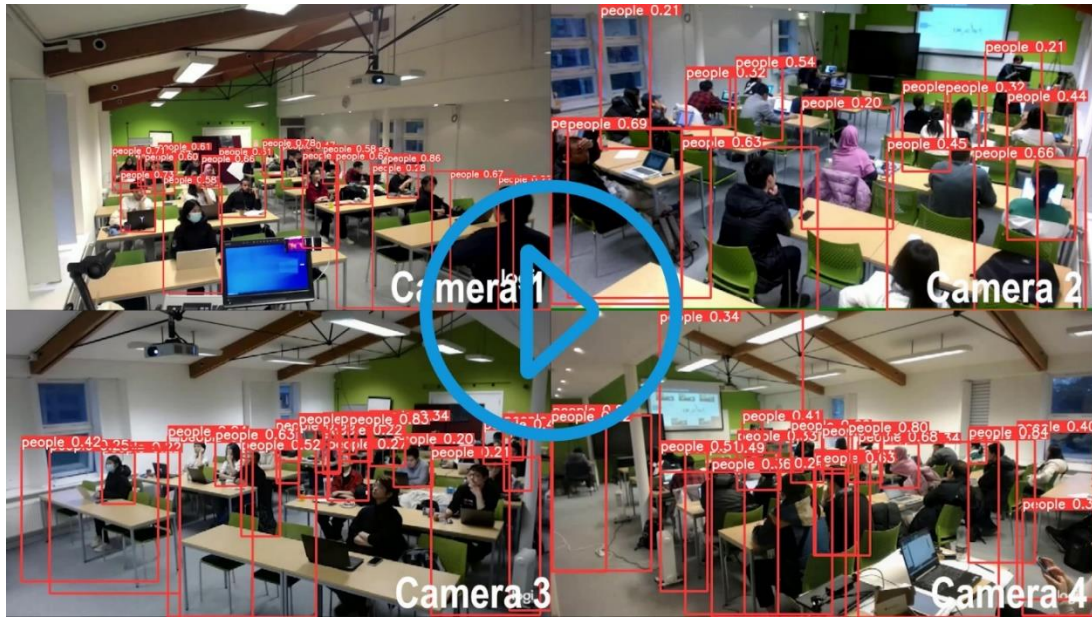


Figure 4-11 A frame taken from the YOLOv8n inference detection videos comparing 4 different cameras at 15:45:00.



Video 4-4 The inference detection videos comparing 4 different cameras using YOLOv8n from 15:45:00 to 15:46:00. The playable video is available at <https://ars.els-cdn.com/content/image/1-s2.0-S2352710224029231-mmc4.mp4>

Figure 4-12 compares the detection results from each of the four cameras. Detailed model performance is also shown in Table 4-7. The occupant number detected by cameras 1 and 3 tends to surpass that of cameras 2 and 4. These contrasting outcomes from different cameras underscore the potential uncertainty in detection when relying on a single view. Considering installation costs, enhancing the accuracy of the deep learning model emerges as a more efficient alternative compared to deploying multiple cameras within a room. Investing in more advanced algorithms can potentially reduce the need for extensive hardware setups, leading to cost savings and simpler implementations.

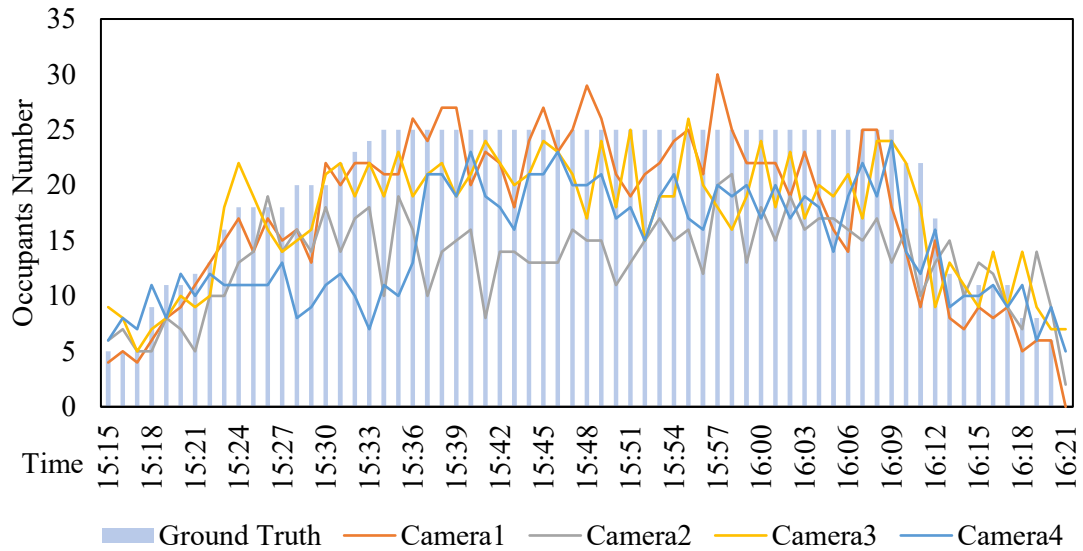


Figure 4-12 Occupancy profiles predicted by the deep learning models based on the different detection camera locations using YOLOv8n compared to ground truth.

Table 4-7 The model performance across all four cameras.

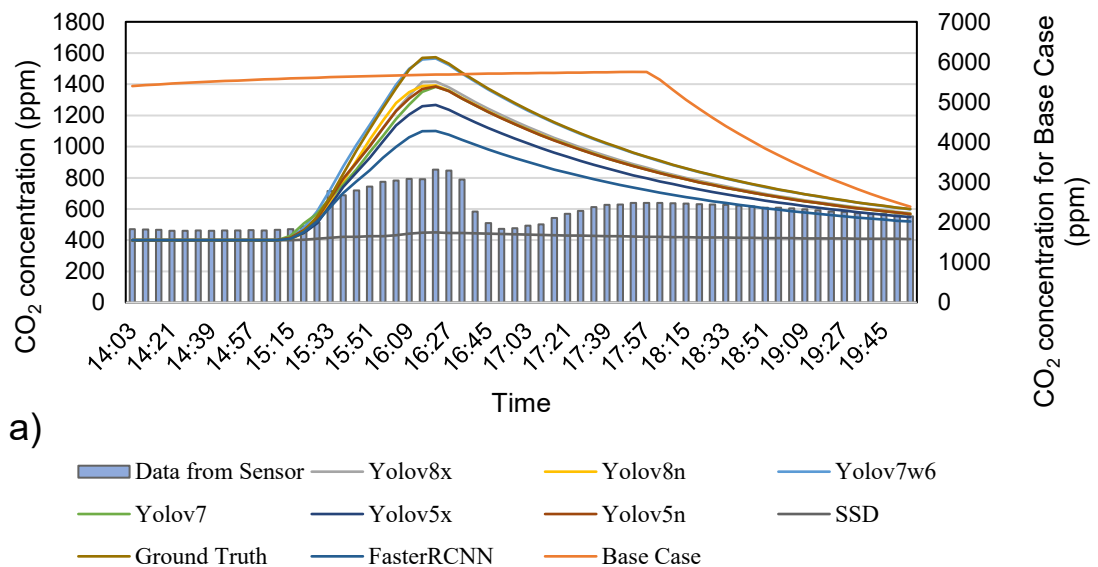
| Metric | Camera 1 | Camera 2 | Camera 3 | Camera 4 |
|-----------|----------|----------|----------|----------|
| Accuracy | 0.72 | 0.51 | 0.69 | 0.65 |
| Recall | 0.78 | 0.71 | 0.90 | 0.82 |
| Precision | 0.90 | 0.64 | 0.74 | 0.76 |
| F1 score | 0.84 | 0.67 | 0.81 | 0.79 |

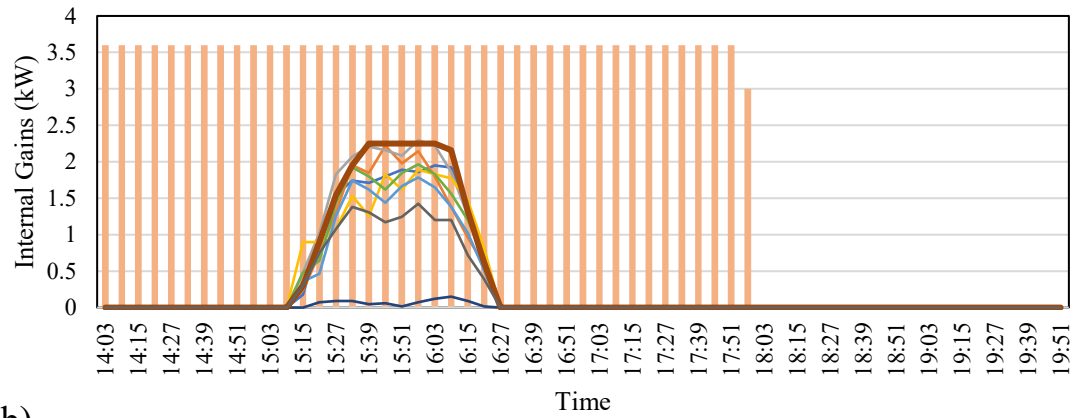
4.3.3 Energy and CO₂ simulation results

An IES VE model was used to simulate the case study building [64] to determine how the occupant detection approach would affect the building's predicted performance, with a focus on energy consumption and CO₂ concentration. Even though the SSD model was wildly off and ill-suited for the application needed, it is nevertheless assessed here to demonstrate its impact on the forecasts. The occupancy profiles generated, encompassing both true and false positive results from the deep learning models, were integrated into the simulation. These profiles were formatted into 5-minute intervals to align with the constraints of the building energy software.

A conventional occupancy profile, designating full occupancy (48 individuals) from 8:00 to 18:00 on weekdays, was established as the "Base Case" for comparison. Figure 4-13a shows the CO₂ concentration trends predicted using the occupancy profiles based on the eight deep-learning models compared with the recorded CO₂ sensor data during the experiment. The CO₂ concentration trend predicted using detections from the deep learning models aligned reasonably well with the recorded data, showcasing a prompt response to occupancy changes from 15:15 onwards, although it generally overpredicted the measurements. This could be an issue with the IES VE modelling, as the ground truth (actual occupancy) also showed discrepancies. In contrast, the "Base Case" simulation exhibited a pronounced discrepancy with the actual data, and the SSD model profile failed to capture the trend.

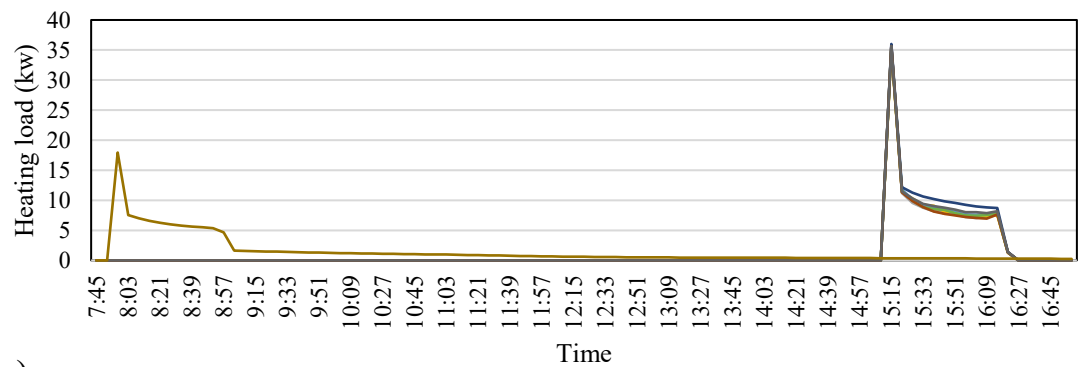
A notable lag of approximately 15 minutes was observed in the recorded CO₂ data from the sensor as occupants began entering at 15:15, with the sensor reflecting changes only around 15:30. This delay is typical for CO₂ sensors, highlighting a temporal limitation in capturing occupancy variations (Liang et al., 2024). The vision-based deep learning approach demonstrated a promising capacity to mitigate this lag, signifying its potential to enhance real-time responsiveness in monitoring and controlling building environments.





b)

Base Case Yolov8x Yolov8n Yolov7w6 Yolov7
Yolov5x Yolov5n SSD Ground Truth FasterRCNN



c)

Yolov8x Yolov8n Yolov7w6 Yolov7 Yolov5x
Yolov5n SSD Ground Truth FasterRCNN Base Case

Figure 4-13 Predicted vs. recorded data for a) CO₂ concentration, b) internal gains, and c) heating loads.

Figure 4-13b displays the internal gains profiles predicted based on the deep learning models, compared to the "Base Case" profile, which assumes full occupancy from 8:00 to 18:00. The actual observed "Ground Truth" profile is highlighted. There is a significant gap between the actual occupancy profiles and the fully occupied profiles commonly used (Azar and Menassa, 2012). Most deep learning models in this study can identify and locate people in the case study room, demonstrating the potential for improving model performance and energy demand prediction accuracy.

Figure 4-13c displays the heating energy load results in the modelled room from 8:00 to 17:00. The "Base Case" simulation showed a significant discrepancy from the "Ground Truth" simulation, indicating that the conventional fixed profile can be inaccurate in certain scenarios. At the beginning of the lecture during the heating period at 15:15, a substantial amount of heating energy was required to maintain the room at the setpoint temperature of 21 °C for the duration of the occupancy period. This need rapidly decreased as occupants entered the room and generated internal heat gains. Compared to the actual profile ("Ground Truth"), the YOLOv8n and YOLOv8x models achieved the most accurate heating energy load predictions, while SSD showed the worst performance.

Figure 4-14 illustrates the predicted heating energy consumption in the case study room on the test day, comparing different models, the "Ground Truth" and the "Base Case" profiles. The results reveal a 13.45% discrepancy between the conventional "Base Case" profile and the "Ground Truth" heating energy consumption. In contrast, the deep learning models show a narrower variation, ranging between 0 and 6.72% from the "Ground Truth" results, except for the SSD model, which barely detects any occupants. The conventional fixed profile sets the heater on continuously, even when there are no occupants, leading to higher energy consumption. The data highlights the potential of deep learning models to accurately capture occupancy changes, improving energy consumption predictions compared to traditional methods. These findings emphasise the shift from static to real-time occupancy detection models. Deep learning models reduce variations in consumption, providing a more reliable foundation for energy management

decisions.

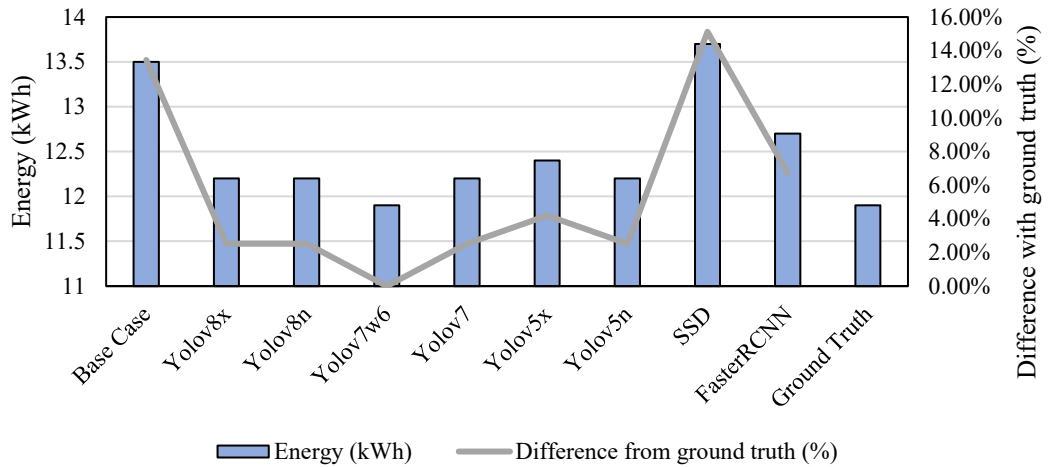


Figure 4-14 The predicted heating energy of the case study room on Dec 2nd, 2022, based on the simulation of deep learning model profiles, the “Ground Truth” profile and the “Base case” profile.

Generally, the deep learning models showed good performance, particularly when compared to the conventional fixed occupancy profile regarding energy and CO₂ predictions. These models were adept at swiftly capturing occupancy variations, faster than the CO₂ sensor, indicating promising avenues for future developments. Although the configuration of the model influenced detection performance, other elements like appliances, furniture, lighting conditions and obstructions also played a role. These factors contributed to the observed changes in predictions and detection performance over the detection period. In this context, YOLOv8x emerged as the most proficient in detection performance, delivering the most accurate predictions on occupant profiles, while YOLOv8n excelled in speed and maintained good accuracy overall.

The vision-based method does present certain limitations, with obstructions being a notable challenge due to its inherent nature. This study evaluated the impact of the position and angle of the cameras on detection performance. The effectiveness of the vision-based method is constrained by the camera's field of view, necessitating strategic

placement and perspective adjustments. Additionally, camera resolution could potentially hinder model performance, especially in detecting minute movements or small objects.

4.4 Summary

A vision-based deep learning approach for real-time occupancy detection in crowded environments is presented in this chapter, with a specific focus on a lecture room at a university. We evaluated eight deep learning models, including SSD, Faster R-CNN, YOLOv5n, YOLOv5x, YOLOv7, YOLOv7w6, YOLOv8n, and YOLOv8x, using a self-compiled dataset during a university lecture experiment. The performance of these models was evaluated in terms of speed of detection, computational requirement, and complexity of the scene. The evaluation revealed varying performance levels among the models. YOLOv8x emerged as the most accurate, with an overall accuracy of 0.77 and an F1 score of 0.87, albeit with a longer inference time. YOLOv8n also demonstrated commendable speed in both training and inference phases while maintaining good accuracy, making it a suitable choice for scenarios prioritising both speed and accuracy. The SSD model, on the other hand, trailed behind significantly, showing a subpar detection ability, particularly struggling to identify occupants unless they were near the camera.

Additionally, this chapter explored the impact of the location and angle of the camera, to assess occlusion challenges often encountered in single-view setups. Specifically, the experiment demonstrated that when one camera missed certain occupants due to obstructions or limited field of view, other cameras positioned at different angles could successfully detect those individuals. For example, individuals missed by camera 2 were detected by cameras 1 and 3, illustrating how a multi-camera setup can compensate for the limitations of a single viewpoint. However, implementing such a multi-camera

system would also increase costs and complexity, necessitating more extensive infrastructure and maintenance.

Examining the impact on the predicted energy consumption, our findings revealed a substantial daily heating energy demand difference of approximately 13.45% when comparing the conventional occupancy profile (Base Case) and the Ground Truth (actual occupancy number). In contrast, deep learning models except the SSD model showed much smaller variations, with a maximum difference of 6.72% compared to the “Ground Truth”. This highlights the potential of the approach to reduce the gap between actual and predicted energy consumption and improve precise, demand-driven building management systems. Although the deep learning models generally overpredicted the recorded data, the CO₂ concentration trends they predicted aligned closely with the recorded data, unlike the Base Case profile. This alignment demonstrates the potential of the proposed method not only to improve the accuracy and reliability of energy performance predictions but also to respond more quickly to occupancy changes than CO₂ sensors.

These findings suggest a promising future for demand-driven management systems. Assessing scalability, conducting comparative studies with other technologies, and gathering user feedback could provide comprehensive insights into the practical deployment and effectiveness of the proposed system in real-world settings. However, while algorithmic performance is critical, the choice of sensing technology also plays a vital role in occupancy detection systems. Different types of cameras, such as standard and thermal cameras, offer unique advantages and face challenges in real-world scenarios. The next chapter explores the comparative performance of standard and thermal cameras in diverse settings, evaluating their suitability for privacy-sensitive environments, complex occupancy scenarios, and energy management applications. By linking algorithmic efficiency with sensor selection, the following chapter aims to

provide a holistic understanding of vision-based occupancy detection for smarter, more energy-efficient buildings.

5. OCCUPANCY PREDICTION PERFORMANCE COMPARISON OF STANDARD CAMERA AND THERMOGRAPHIC IMAGING

5.1 Introduction

This chapter presents a comparison study for building occupancy prediction with deep learning methods with both standard cameras and thermal images. Many HVAC systems operate on fixed or predefined schedules, which assume steady occupancy patterns throughout the day. This often results in energy waste in unoccupied or partially occupied spaces (Tien et al., 2020c). Therefore, effective occupant detection and monitoring are essential for optimizing energy performance by enabling dynamic control strategies, improving simulation accuracy, and maintaining indoor environmental quality.

Traditional occupant detection methods, such as passive infrared sensors, CO₂ sensors, and radio frequency identification, are widely used in building control systems (Zhang et al., 2022, Franco and Leccese, 2020). While these methods capture basic occupancy signals like motion or changes in ambient CO₂ concentration, they may offer limited insights into occupant distribution, behaviour, or activities. More advanced approaches, such as systems utilizing Wi-Fi (Alishahi et al., 2022) and Bluetooth signals (Park et al., 2019), have been developed to detect and track occupants with greater precision and contextual awareness, thereby enhancing dynamic building control. Despite their potential, these methods face challenges such as susceptibility to signal interference, significant infrastructure costs, and privacy concerns associated with the continuous tracking of personal devices.

Recent advances in computer vision provide richer contextual information, enabling building control systems to detect presence and understand occupant locations (Hu et al., 2023), movement patterns, and interactions within a space (Aliero et al., 2022). However, vision-based solutions using standard (RGB) cameras present privacy challenges (Winkler and Rinner, 2013), particularly in commercial or sensitive environments, and are prone to false detections caused by indoor elements like portraits and photographs.

Thermal imaging has emerged as a promising alternative to address these limitations. By capturing temperature variations instead of visual details, thermal cameras safeguard privacy by rendering faces and personal identifiers less discernible (Qin et al., 2021). Additionally, thermal cameras are robust to variations in lighting conditions, making them suitable for low-light or nighttime scenarios where standard cameras are less effective. However, the effectiveness of thermal imaging for occupant detection can be influenced by factors such as emissivity settings, ambient temperature fluctuations, and object heat signatures resembling human presence. While the higher cost of thermal cameras has traditionally limited their adoption, the emergence of low-cost and low-resolution thermal cameras offers more accessible options (Metwaly et al., 2019a), although with trade-offs in image quality and detection accuracy (Kraft et al., 2021, Sirmacek and Riveiro, 2020).

Previous works (Cosma and Simha, 2018) have successfully utilized thermal imaging and computer vision to assess thermal comfort. However, few studies (Acquaah et al., 2021) have focused on using thermal imaging for indoor occupancy detection, especially in real-time applications employing algorithms like single-shot detectors such as YOLO (Long et al., 2020), which provide faster and more efficient processing compared to older models such as traditional machine learning models (e.g., k-means, SVM, RF, and GNB) (Sahoo and Lone, 2023), classic neural networks (e.g., MLP) (Long et al., 2020), and earlier deep learning architectures (e.g., ResNet-50 and VGG-

16) (Acquaah et al., 2021). YOLO's architecture enables simultaneous object detection and classification in a single step, reducing computational demands and improving its real-time applicability for dynamic and complex indoor environments.

Furthermore, most studies (Figure 5-1) rely on ceiling-mounted, top-view and low-resolution cameras (Sahoo and Lone, 2023) that capture heads rather than full bodies, which could potentially restrict its ability to analyse occupant behaviour or activities (for example using appliances, opening windows). Moreover, comparisons between low-cost thermal cameras and standard vision-based occupancy detection systems are lacking. Research evaluating these technologies in real environments is also limited. Additionally, no studies have systematically examined the impact of various heat sources, such as TV screens, computers, heated seats, or coffee cups, as potential sources of error for thermal imaging (Figure 5-1). For standard cameras, challenges like detecting pictures or screens displaying people in rooms also remain underexplored.

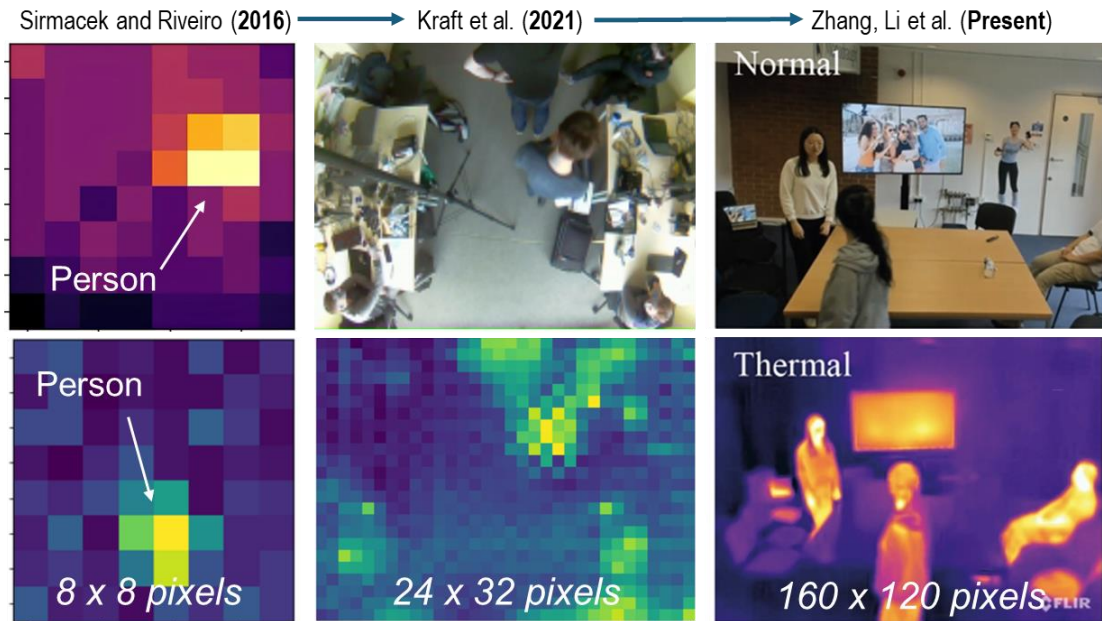


Figure 5-1 Evolution of the use of thermal imaging across occupancy studies, from low to medium pixel grids (Kraft et al., 2021, Sirmacek and Riveiro, 2020) to higher detail (present).

Building on this context, the study explores real-time occupancy prediction performance using low-cost thermal cameras and standard cameras across a range of indoor settings, employing advanced deep-learning algorithms. Specifically, it utilizes the YOLOv8/v10 models to detect occupants under different experimental conditions, including Same-Video, Split-Video, and Cross-Video setups (Figure 5-2). These experiments also emphasize the importance of creating robust datasets, particularly for thermal image detection. Creating such datasets involves carefully selecting environments with varying levels of complexity, incorporating diverse heat sources, and ensuring representative scenarios to account for potential sources of error, such as reflections and heat signatures from objects like coffee cups or electronic devices. These experiments are carefully designed to test how well the models adapt to varying environmental complexities and occupant densities, offering a thorough evaluation of their practical effectiveness in real-world scenarios while addressing key research gaps from prior studies.

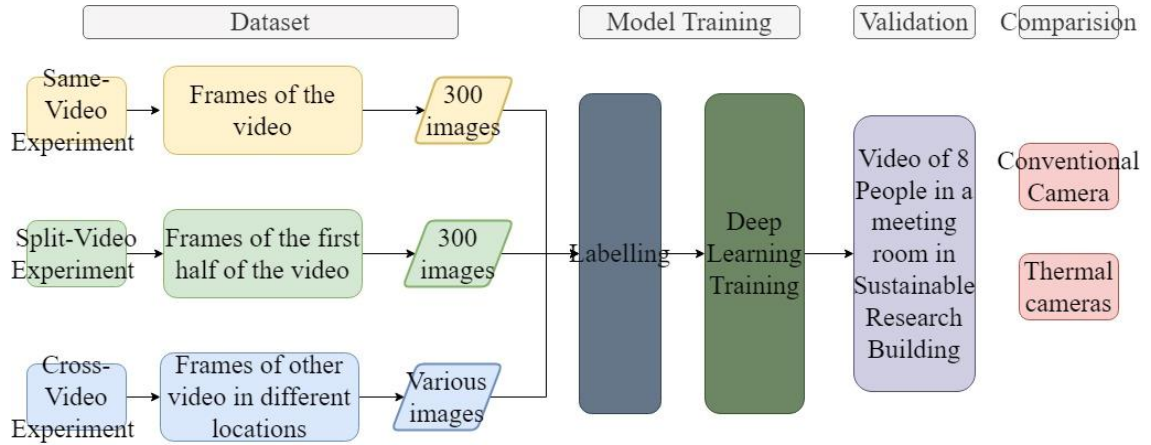


Figure 5-2 The workflow for the comparison of standard and thermal cameras in vision-based occupancy detection.

5.2 Occupancy prediction methodology

This chapter investigates the performance of standard and thermal cameras for occupancy prediction through experiments designed to explore datasets, model training, validation, and comparative analysis. There are three different kinds of experimental setups in this research namely, Same-Video Experiment, Split-Video Experiment, and Cross-Video Experiment. For all experiments, both standard camera and thermal camera videos are recorded. The Same-Video Experiment extracted frames from a video captured in a meeting room with 8 occupants which serves as both the training dataset and validate it on the same video while expected to have high accuracy and serves as a baseline for comparison rather than reflecting real-world applicability. In the Split-Video Experiment, the first half of the video is used for dataset generation and the second half for validation, which is to test in a realistic scenario where pre-data collection in the same environment enables model training. Finally, the Cross-Video Experiment, which uses datasets from other videos and validates the model in different videos captured in separate locations, aims to evaluate the model's generalization capability. Various settings of experiments were conducted, and the details will be discussed later in this section.

5.2.1 Case study experiment setups

This chapter evaluates the performance of standard and thermal cameras for occupancy prediction through four field studies, all conducted at the University of Nottingham, UK. In each field study, a thermal camera (FLIR ONE Pro) and a standard webcam (Logitech C920) were positioned and recorded videos at the same time as shown in Figure 5-3. These cameras were selected to represent different technologies, with the FLIR ONE Pro capturing thermal images and the Logitech C920 recording standard video. The Logitech C920 webcam was chosen for its capability to record full-HD 1080p video at 30 frames per second (fps) and its 78-degree field of view , which provided a detailed and comprehensive perspective of the test environment. The high resolution and wide field of view make it an effective tool for detecting and analysing occupant movements.

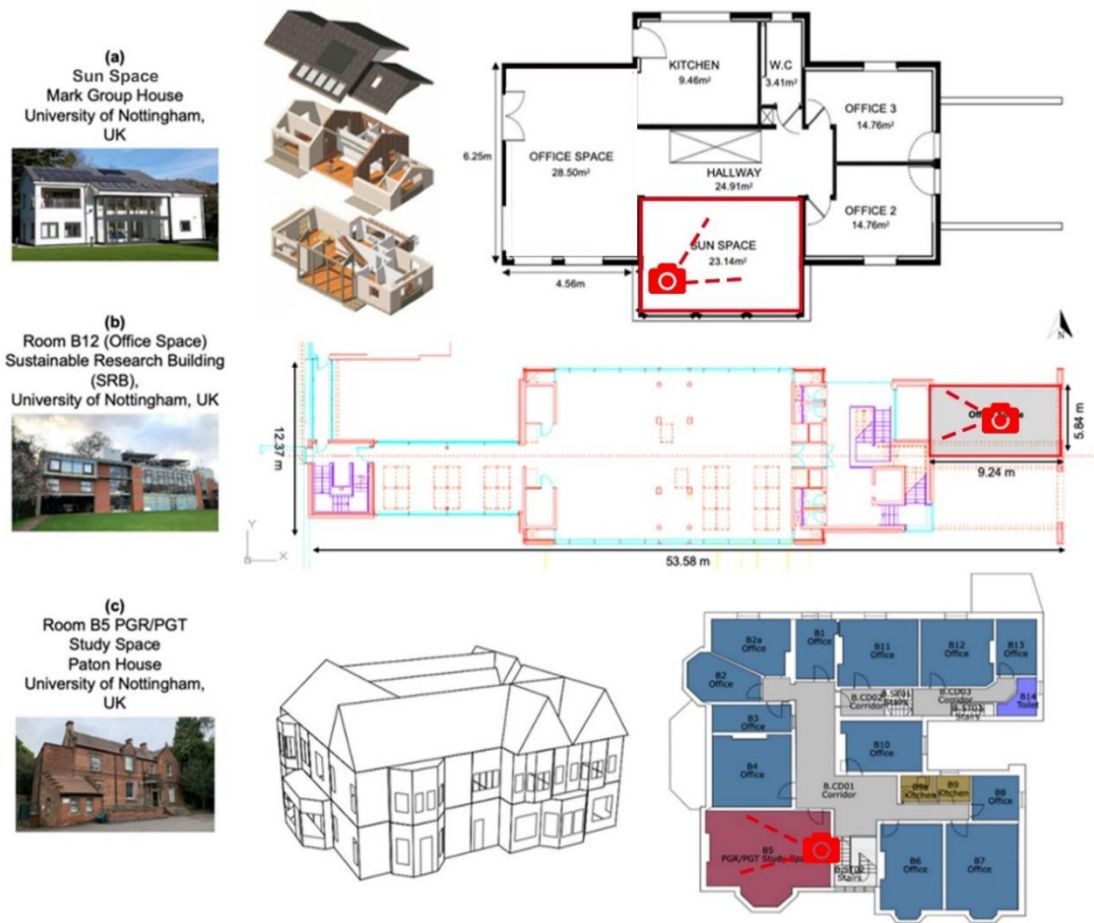


Figure 5-3 The details of the buildings where the field studies are conducted.

The FLIR ONE Pro thermal camera was selected for its affordability and suitability for building applications, costing approximately £300–400. Despite its relatively low resolution of 160×120 , the thermal camera ensures privacy by capturing only temperature variations and patterns, avoiding identifiable visual details. It operates within a spectral range of 8–14 μm with a horizontal field of view of $50^\circ \pm 1^\circ$. The differing fields of view between standard and thermal cameras need adjustments during comparison. To ensure consistency, the images captured by the Logitech C920 were cropped to match the narrower field of view of the FLIR ONE Pro.

The accuracy of the FLIR ONE Pro thermal camera is $\pm 3^\circ\text{C}$ or $\pm 5\%$, and it has an object temperature range of -20°C to 120°C . To optimize its performance for human detection, the emissivity was set to 0.98, which is appropriate for human skin, and the reflection

temperature was set to the factory default of 22°C. The FLIR ONE Pro is factory-calibrated and performs automatic flat-field correction. During this process, the shutter closes, and auto-calibration is executed every 3 minutes to account for changes in the thermal scene. The infrared (IR) scale was set to automatic for all experiments to ensure consistency in thermal image quality. The thermal camera recorded videos of occupants' movements, which were then converted into individual frames to create datasets for training the deep learning models.

The settings for each field study are detailed in Table 5-1. Field study 1 was conducted in a small room in the Mark Group House, involving three occupants. The scenario was considered simple, with occupants seated separately most of the time, providing a controlled and basic setting to establish baseline performance for both standard and thermal cameras. Field study 2 took place in a larger room in Paton House, involving seven occupants. In this field study, occupants sometimes sat close together, offering a more complex and dynamic environment to test the cameras' ability to differentiate individuals at closer distances. Field study 3 was conducted in a meeting room in the Sustainable Research Building, involving eight occupants. This test introduced additional complexities, such as posters of human figures on the walls and people's images displayed on a TV screen. These additional elements were intentionally included to evaluate the performance of both standard and thermal cameras for specific challenges, such as distinguishing between real occupants and distractions within the environment. Field study 4 served as an additional test and involved capturing random thermal images of students in various classrooms across the campus. This test captured scenarios where students were seated, leaving their chairs, or interacting with screens.

Table 5-1 Details of the setting of experiments conducted for this study.

| | Location | Occupancy Number | Area (m²) | Video Duration | Description |
|---------------|-----------------------------|-------------------------|-----------------------------|-----------------------|--|
| Field study 1 | Sun Space, Mark Group House | 3 | 23.14 | 35 mins | A meeting room with max 3 sitting occupants. |

| | | | | | |
|---------------|---|---------|-------|---------|--|
| Field study 2 | Room B5, Paton House | 7 | 36.62 | 38 mins | A meeting room with occupants, from an empty room to a crowded scenario. |
| Field study 3 | Room B12, Sustainable Research Building | 8 | 53.96 | 10mins | A meeting room with occupants sitting and walking, with a TV and a poster in the background. |
| Field study 4 | Random Classrooms | Various | - | - | Random occupants in different classrooms. |

5.2.2 Training dataset generation

For the Same-Video Experiment and Split-Video Experiment, both the training dataset and validation video were derived from Experiment 3, conducted in the Sustainable Research Building. This experiment involved a 10-minute recording that captured a dynamic scenario with people entering the room, walking around, and sitting together. Special features of the environment included a TV screen displaying images of people and a poster of human figures on the wall. These elements were intentionally included to test the deep learning model's ability to distinguish between real occupants and visual distractions, such as images of people on screens or posters.

The Same-Video Experiment used the entire video from Experiment 3 for both the dataset and validation, aiming to assess the model's peak performance under ideal but unrealistic conditions. 300 frames in the video were taken and labelled as the training dataset. This setup, though unlikely to reflect real-world applications, serves as a benchmark for comparison with other experiments.

The Split-Video Experiment took 300 frames in the first half of the Experiment 3 video for the training dataset and the remaining half video for validation. This configuration used pre-collected data from the same room for training. It evaluates the model's adaptability to slight variations within the same environment while still maintaining continuity between training and validation datasets. This experiment bridges the gap

between ideal and real-world applications, providing insights into the model's performance in more achievable conditions.

For Cross-Video Experiments, datasets were created using images captured in different locations to evaluate the model's performance in more generalised and diverse situations. This approach was designed to test the adaptability of the deep learning model beyond a single environment and examine its effectiveness in scenarios that closely resemble real-world applications. Three datasets were developed, each representing different levels of complexity, ranging from basic scenarios with a small number of people to crowded scenarios with movement and overlapping individuals. The aim was to assess how variations in environmental settings and occupant density impact the model's accuracy. The details of the datasets, including the characteristics of each scenario, are outlined in Table 5-2.

Table 5-2 The details of the different datasets used as Cross-video experiments in this study.

| | Source | Images | Description | Aims | Dataset |
|-----------|-------------------|--------|--|---|---|
| Dataset 1 | Field study 1 | 300 | Contains basic scenario | Set a baseline | Standard: https://app.roboflow.com/ds/ZWRk34k5L8?key=W7OA90qchf Thermal: https://app.roboflow.com/ds/ByXqPWkzwc?key=KEOuoOiQs4 |
| Dataset 2 | Field study 1+4 | 785 | The basic scenario with additional images of random people | Increase the dataset and add images of people in different background | Standard: https://app.roboflow.com/ds/yFGSO7dnBE?key=vs1aLGOqgD Thermal: https://app.roboflow.com/ds/Ge1nl15j5h?key=kjI3Dgyph7 |
| Dataset 3 | Field study 1+2+4 | 985 | The basic scenario and crowded scenario | Increase dataset with crowded scenario | Standard: https://app.roboflow.com/ds/95n4NmI7Pq?key=y2Jz6qzs2O |

| | | | | | |
|--|--|--|--|--|---|
| | | | | | Thermal: https://app.roboflow.com/ds/GONHsm1UjG?key=BBBQpCZCUM |
|--|--|--|--|--|---|

In this study, all datasets were created using frames extracted from the field study videos, which were then manually annotated. The images were randomly divided into three subsets for training, validation, and testing, with proportions of 70%, 20%, and 10%, respectively. All datasets are publicly available on Roboflow, an open-source platform for sharing datasets (Jocher, 2023). The annotation process was carried out using Roboflow’s tools, which generate bounding boxes around each occupant in the images. The annotations were saved in YOLO text format, ensuring compatibility with the machine learning frameworks. Each frame was carefully reviewed and annotated manually to maintain high data quality for training and evaluating the deep learning models. The annotation workflow included loading images, creating bounding boxes, assigning labels, and exporting the annotations in the required format, ensuring the datasets were well-prepared for further model development.

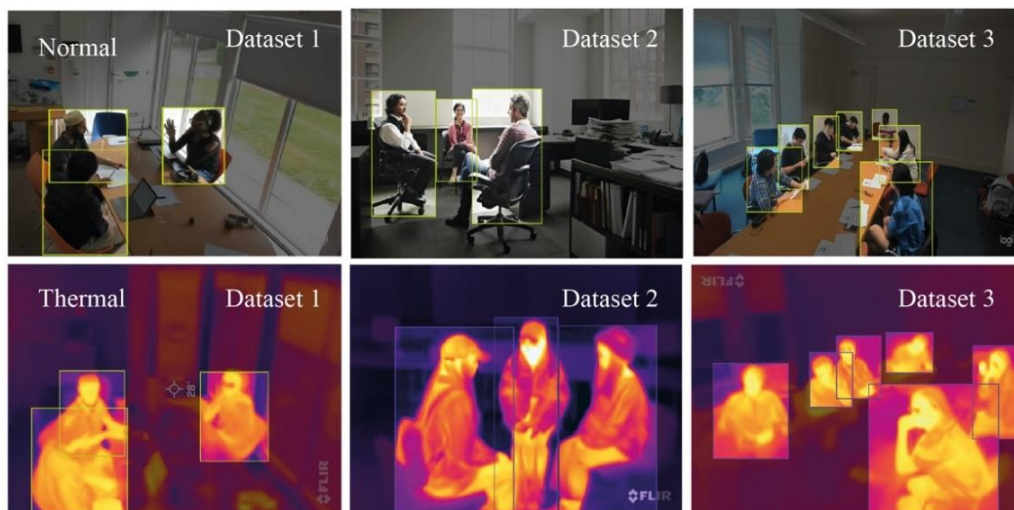


Figure 5-4 Example images from datasets 1, 2 and 3. Dataset 1 represents simple scenarios and 2 and 3 for crowded ones.

Each image was annotated with bounding boxes specifying the exact location of humans, using coordinates for x_center , y_center , width, and height. The annotations were saved

in.xml files for YOLO input. Preprocessing steps included resizing images to 640x640 pixels and applying auto-orientation to ensure compatibility with the models, preventing memory leaks, poor performance, and imprecise results. Figure 5-4 shows examples of standard and thermal images in Datasets 1, 2 and 3.

5.2.3 Training model

YOLOv8 (Jocher, 2023) and YOLOv10 (Wang et al., 2024) are selected for real-time deep-learning occupancy detection, as they represent the latest advancements in the YOLO series. These models have demonstrated good performance in terms of both accuracy and speed, making them ideal for dynamic and high-density environments. Compared to earlier YOLO versions, YOLOv8 and YOLOv10 offer improved detection accuracy, particularly for smaller objects, which is critical for distinguishing individuals in scenarios with overlapping occupants. Additionally, the model's ability to perform high-speed inference makes them well-suited for real-time applications.

The YOLO models were trained using the Pytorch framework in Google Colab (Bisong and Bisong, 2019), which provides free access to an NVIDIA T4 Tensor Core GPU. This GPU features 2560 CUDA cores, a 1590 MHz graphics clock, 320 GB/s memory bandwidth, and 15 GB of GPU memory, enabling efficient processing for large-scale datasets. The training process was set to run for 300 epochs, and the training was automatically terminated if the mean Average Precision (mAP) did not improve over 100 epochs, ensuring efficient use of computational resources and preventing overfitting. The details of the training process for different tests conducted in this study are presented in Table 5-3. And

Figure 5-5 demonstrates the normalized confusion matrices for all experiments. All tests were conducted using both standard and thermal images and videos. YOLOv8 was employed for most experiments, except for the final test, which used YOLOv10 to compare its performance with the latest YOLO model with the best performance dataset

in previous tests. During training, the validation subset was used to monitor the model's performance, with key metrics such as mAP@50, precision, and recall evaluated after each epoch. This monitoring ensured the early detection of overfitting and informed the fine-tuning of hyperparameters.

Table 5-3 The training details of all deep learning tests in this chapter.

| Exp. | Training dataset | Standard/ Thermal | Model | Training time (Hour) | Epochs | Precision | Recall | mAP ₅₀ | mAP ₀₋₉₅ ⁵ |
|-------------|--------------------------|----------------------|---------|--|--------|-----------|--------|-------------------|----------------------------------|
| | | | | Low  High | | | | | |
| Same-Video | Experiment 3 | N | YOLOv8 | 0.60 | 300 | 0.97 | 0.94 | 0.97 | 0.8 |
| | | T | | 0.59 | 300 | 0.97 | 0.96 | 0.98 | 0.78 |
| Split-Video | The first half of Exp. 3 | N | YOLOv8 | 0.60 | 300 | 0.99 | 0.98 | 0.99 | 0.92 |
| | | T | | 0.58 | 300 | 0.97 | 0.95 | 0.98 | 0.81 |
| Cross-Video | Dataset 1 | N | YOLOv8 | 0.70 | 300 | 0.95 | 0.92 | 0.98 | 0.9 |
| | | T | | 0.63 | 300 | 0.99 | 0.99 | 0.99 | 0.94 |
| | Dataset 2 | N | YOLOv8 | 1.36 | 271 | 0.94 | 0.88 | 0.94 | 0.73 |
| | | T | | 1.17 | 250 | 0.82 | 0.71 | 0.78 | 0.44 |
| | Dataset 3 | N | YOLOv8 | 1.84 | 300 | 0.93 | 0.92 | 0.95 | 0.73 |
| | | T | | 1.83 | 281 | 0.87 | 0.78 | 0.84 | 0.51 |
| | Dataset 3 | N | YOLOv10 | 1.92 | 249 | 0.93 | 0.91 | 0.94 | 0.7 |
| | | T | | 0.51 | 66 | 0.77 | 0.71 | 0.78 | 0.43 |

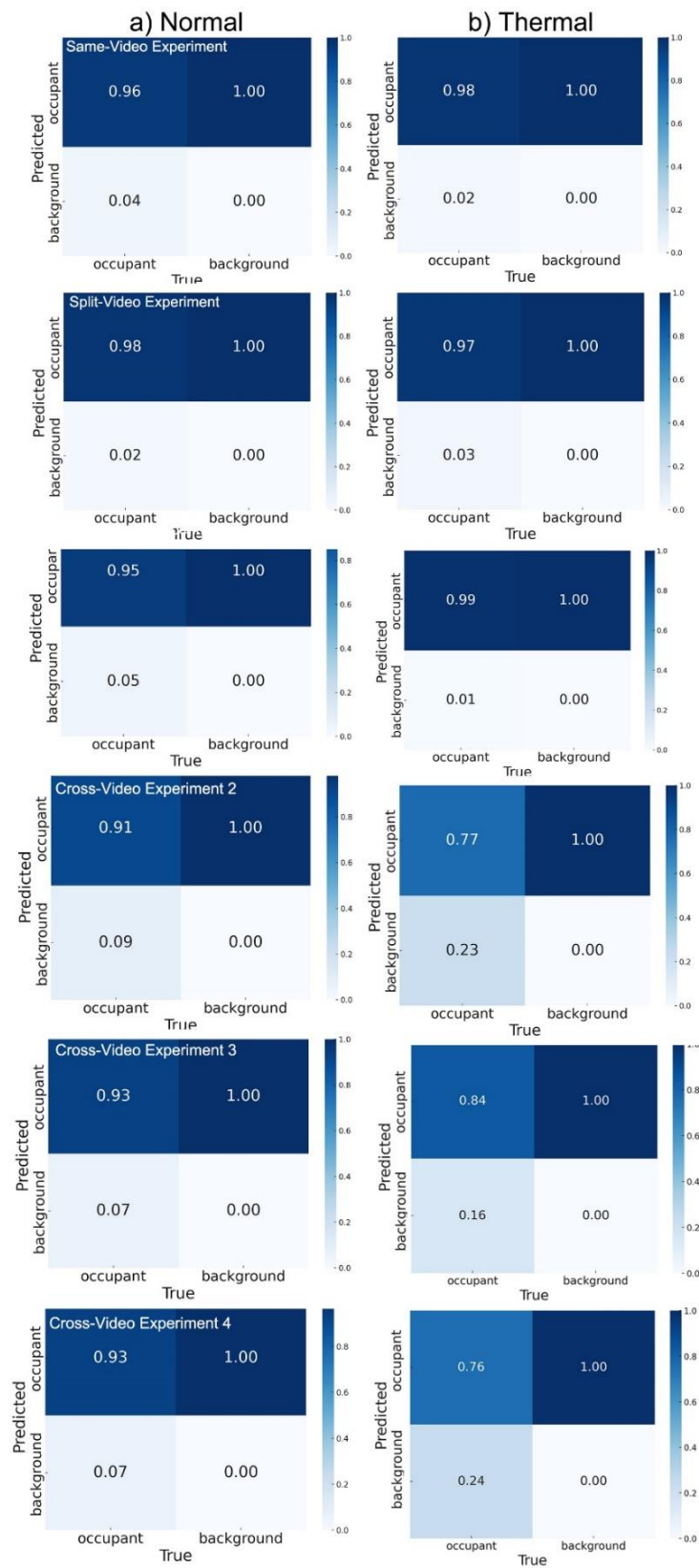


Figure 5-5 The normalized confusion matrix for the Same-Video Experiment, Split-Video Experiment and Cross-Video Experiments for (a) standard and (b) thermal datasets.

For the Same-Video Experiment, both standard and thermal tests achieved high mAP@50 values, reflecting good accuracy when trained and validated on identical datasets. This result highlights the models' capability to fit and perform well in controlled conditions, although it may not fully reflect real-world generalization. In the Split-Video Experiment, YOLOv8 demonstrated good accuracy, with the standard dataset achieving the highest mAP@50 and mAP@50-95 scores among all setups. This outcome indicates the model's ability to adapt effectively to slight variations within the same environment. The high performance of YOLOv8 in this scenario suggests that it is well-suited for situations where pre-collected data from the same location can be used for training.

The Cross-Video Experiments demonstrated the models' generalization capabilities when datasets were collected from different locations. The results reveal a decline in mAP@50-95 scores for thermal datasets, particularly in Dataset 2 and Dataset 3. This indicates that thermal datasets, while effective in controlled environments, are more sensitive to changes in environmental context and occupant behaviours. For instance, in Dataset 2 and Dataset 3, where the environments were more dynamic and occupant interactions were more complex, the thermal models struggled to maintain high recall and precision compared to their performance in Dataset 1 which has a simpler setting. In contrast, the standard camera datasets showed more stable performance in all cross-video experiments. This stability can be attributed to the high spatial resolution and detailed visual data provided by standard cameras, enabling the models to better distinguish individual occupants and adapt to environmental variability. However, the dependency on adequate lighting remains a limitation for standard cameras, making them less effective in low-light or privacy-sensitive scenarios.

5.3 Experiment Results

5.3.1 Video inference results

In addition to the validation process provided by the deep learning algorithm during training, this study employed the video from Field Study 3 as an inference video to evaluate the models' performance under real-world conditions. During inference, the video was sliced into single frames by the algorithm, and each frame was processed to generate results indicating the number of occupants to form an occupancy profile in the room. To validate the model's accuracy, the generated occupancy profile was compared with manual occupant counts from the same video which served as the ground truth for evaluating the inference results. Since occupant numbers in the experiment do not change rapidly, the deep learning profile generates the occupant count by taking the average over every 10-second interval. This interval was chosen to ensure consistency and reduce noise in the data, while accurately capturing occupancy patterns.

In terms of the value for the ground truth, the average number of occupants in each 10s was manually counted. In general, the number of occupants does not change in an interval but in some cases, people would enter or leave the video at just the end of one interval or the start of the next interval. Since it was difficult to determine the number variation between two intervals, the 10s were further split into two 5s. That is, when the occupancy number in the first 5s is x , and in the later 5s is $(x \pm 1)$, the ground truth number would be x . If the variation of occupancy number were larger, e.g. $\geq (x \pm 2)$, the average of these two 5s would be calculated.

To obtain an accurate ground truth number, we counted the number from the video generated by a standard camera, as the occupants were easier to recognise. The accuracy could be calculated as:

$$Accuracy = \left(1 - \frac{ABS(Detect - GroundTr)}{GroundTr} \right) * 100 \quad (4 - 1)$$

Where $ABS()$ is the function of absolute value; $GroundTr$ is the ground truth value of occupancy every 10 seconds; Detect is the average detected occupancy every 10 seconds. The detailed accuracy results for all experiments in this study are listed in Table 5-4.

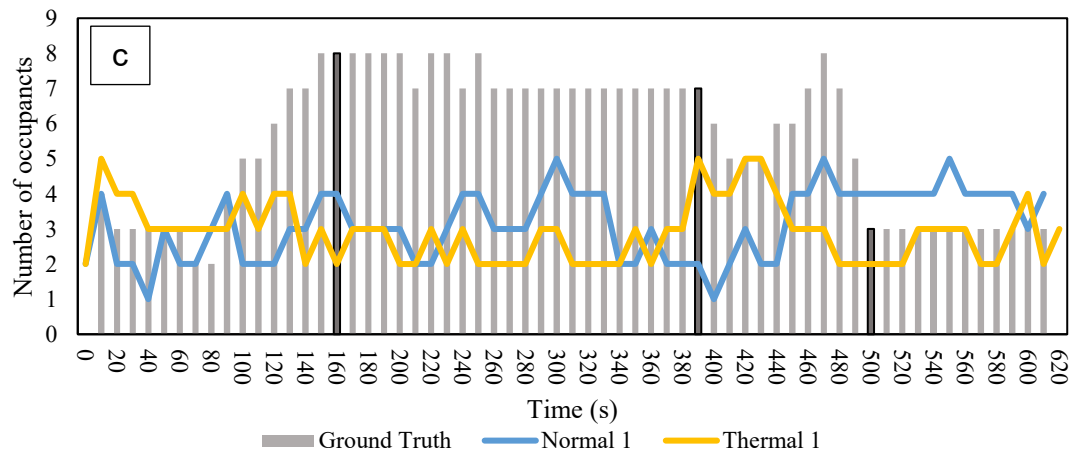
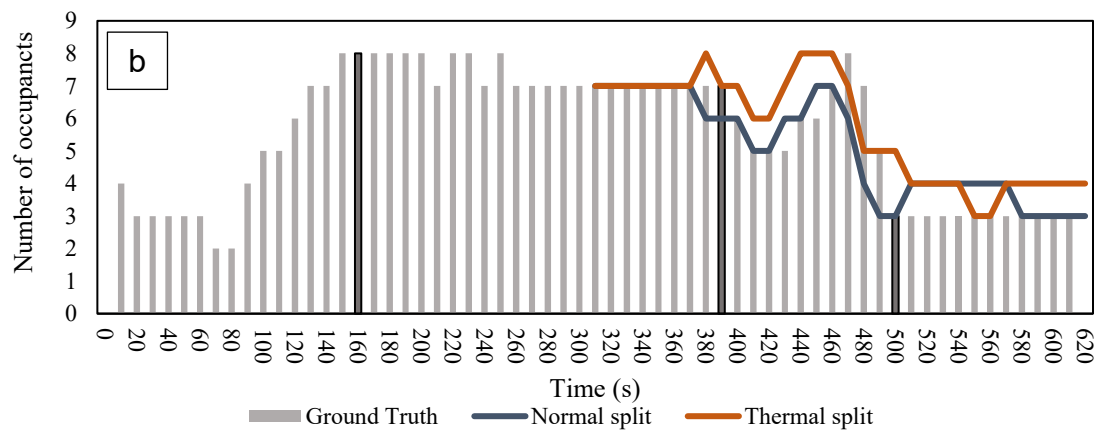
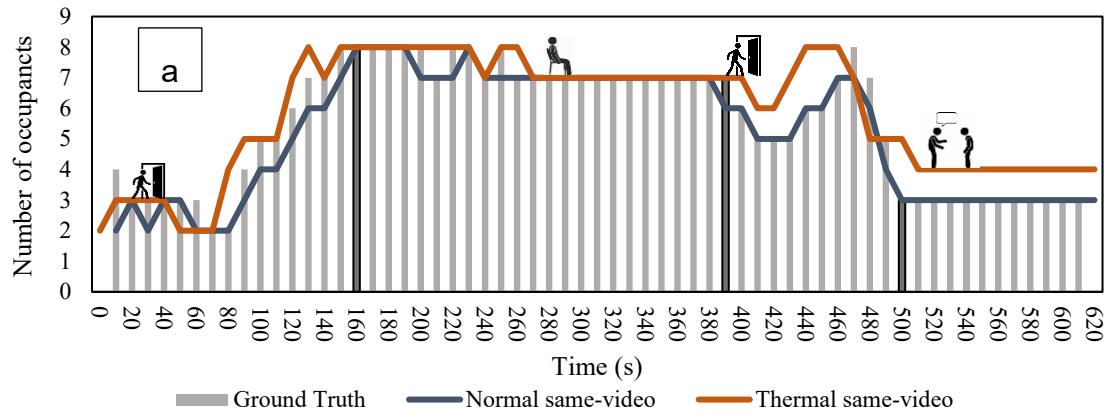
Table 5-4 The detailed accuracy results for all experiments with video inference.

| Experiment | Training dataset | Validation Video | Model | Accuracy (%) | |
|-------------|---------------------------------|------------------|---------|--------------|---------|
| | | | | Low | High |
| | | | | Standard | Thermal |
| Same-Video | Field study 3 | Field study 3 | YOLOv8 | 94 | 93 |
| Split-Video | The first half of Field Study 3 | Field study 3 | YOLOv8 | 90 | 91 |
| Cross-Video | Dataset 1 | Field study 3 | YOLOv8 | 53 | 53 |
| | Dataset 2 | Field study 3 | YOLOv8 | 71 | 71 |
| | Dataset 3 | Field study 3 | YOLOv8 | 78 | 88 |
| | Dataset 3 | Field study 3 | YOLOv10 | 78 | 83 |

The accuracy was determined by comparing the model's predicted occupant counts to the manually established ground truth. YOLOv8 served as the primary model, while YOLOv10 was applied to the dataset with the best performance to evaluate whether further improvements in accuracy could be achieved. The Same-Video Experiment achieved the highest accuracy, with YOLOv8 producing results of 94% for the standard dataset and 93% for the thermal dataset. In the Split-Video Experiment, where the first half of the video was used for training and the second half of the video for validation, accuracy was slightly lower but still strong, with values of 90% and 91% for standard and thermal datasets, respectively. The highest accuracy observed in this experiment serves as a benchmark for the deep learning model, offering a basis for comparison with other experimental setups and establishing the expected upper bounds of model performance under controlled scenarios.

The Cross-Video Experiments presented a more challenging evaluation with datasets collected in different locations for training. Dataset 1 showed the lowest accuracy, at 53% for the standard and thermal datasets. The accuracy was improved by adding Dataset 2, where both standard and thermal datasets achieved 71%. Dataset 3 showed the best results among the cross-video tests, with YOLOv8 achieving 78% accuracy for the standard dataset and 88% for the thermal dataset. Since the highest accuracy was generated from Dataset 3, the cross-video experiment 4 which applied YOLO v10 was also tested on this dataset and produced comparable results, achieving 78% accuracy for the standard dataset and 83% for the thermal dataset. This demonstrates the potential of the newer YOLO version to deliver competitive performance.

In general, simpler datasets often struggle to perform well when applied to more complex validation videos, as they lack the diversity and variability needed to generalise effectively across different scenarios. This limitation was evident in cases where training datasets derived from straightforward scenarios, such as Dataset 1 in the Cross-Video Experiments, produced lower accuracy when validated on the more intricate video from Field Study 3. However, as the training dataset becomes more complex and diverse, the model's performance improves. This was observed with Dataset 3, where the inclusion of more varied scenarios and environmental conditions during training resulted in higher accuracy during validation.



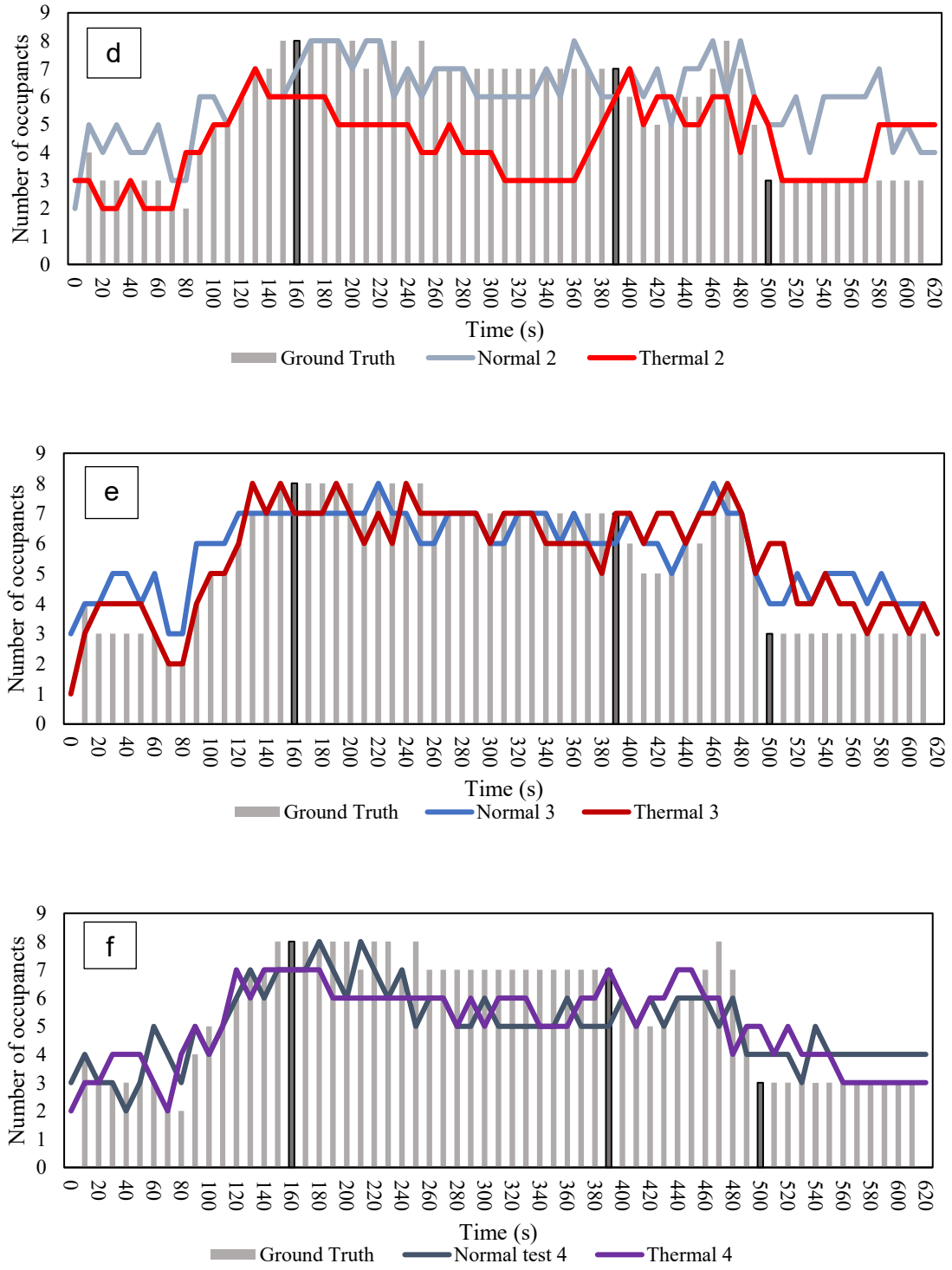


Figure 5-6 The generated occupancy profiles in each experiment. (a) Same video, (b) Split video, and Cross-Video (c) Experiment 1, (d) Experiment 2, (e) Experiment 3, (f) Experiment 4. Three dark grey bars indicate the boundary of four stages of the

inference video, which are entering the room, discussion (sitting), leaving the room, and discussion (standing).

When comparing the performance of standard and thermal datasets, it can be noticed that standard datasets outperformed thermal datasets in most tests. This is likely due to the higher resolution and richer visual details provided by standard datasets, which enable more accurate detection and differentiation of occupants. However, in cross-video validations, thermal datasets exhibited competitive performance and, in some cases, surpassed standard datasets, particularly with Dataset 3. This highlights the potential of thermal cameras in privacy-sensitive applications and under complex or dynamic conditions, where their ability to preserve anonymity while maintaining sufficient accuracy becomes critical. Also, YOLO models were pre-trained on standard image datasets, introducing a bias toward visible-spectrum images. Therefore, the thermal datasets may require additional tuning or pretraining specific to thermal images to unlock their full potential. Such adaptations could enhance the performance of thermal cameras, particularly in scenarios that demand high privacy preservation or operation in low-light environments.

Figure 5-6 shows the result of occupancy profiles in each experiment with a comparison between ground truth and detecting results generated by standard and thermal datasets. In the Same-Video Experiment, the occupancy profiles for both standard and thermal datasets closely align with the ground truth. The Split-Video Experiment shows a similar trend, only the latter half of the video is shown in the occupancy profile, as the first half is used as the training dataset. For the Cross-Video Experiments, the performance of both datasets varies based on the complexity of the training data and the details will be discussed in the next section.

5.3.2 Different scenarios result in Cross-Video Experiments

To further investigate the result in Cross-Video Experiments, the inference video is divided into four stages as shown in Figure 5-7, which are 1) entering the room, 2) discussing (sitting), 3) leaving the room, and 4) discussing (standing). These stages were categorized based on the number of occupants and the extent of occupant overlap, which influenced the complexity of the detection task. The first stage, entering the room, represents a relatively simple scenario with approximately three occupants, each separated within the room. The lack of overlapping individuals and lower occupant density in this stage make it less challenging for the model. In contrast, the second and third stages, which involve discussion while sitting and occupants leaving the room, are more complex. These scenarios feature approximately eight occupants, either sitting in close or moving within the room, increasing overlapping and making detection more difficult. The fourth stage, discussion while standing, is also classified as a crowded scenario due to occupant overlap, which challenges the model's ability to distinguish individuals.

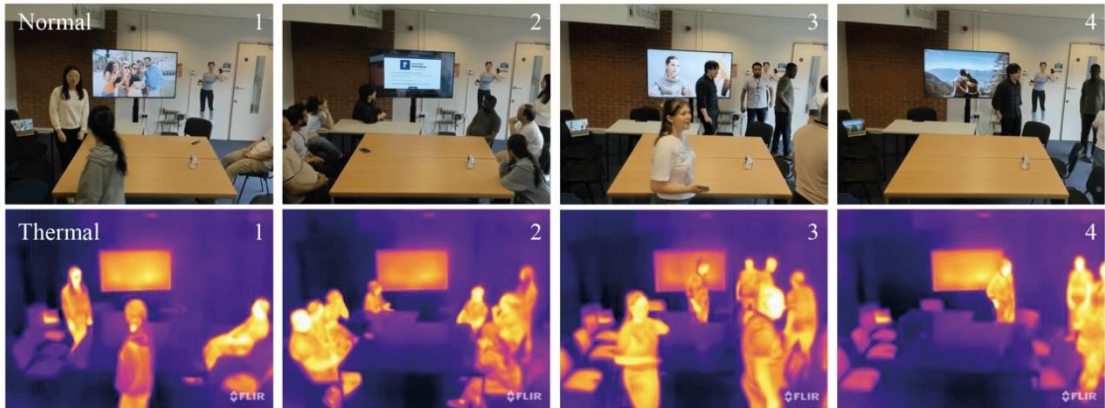


Figure 5-7 The examples of inference video. From Stages 1 to 4: entering the room, discussion (sitting), leaving room, and discussion (standing).

In the entering stage, the number of occupants was around three and gradually increased after 90 seconds. From Experiment 1 to Experiment 3, both the standard and thermal

cameras demonstrated improvements in their ability to detect entering occupants, indicating that adding more training data improved the models' performance. Both cameras were capable of capturing behaviours such as sitting, walking, and standing throughout the experiments.

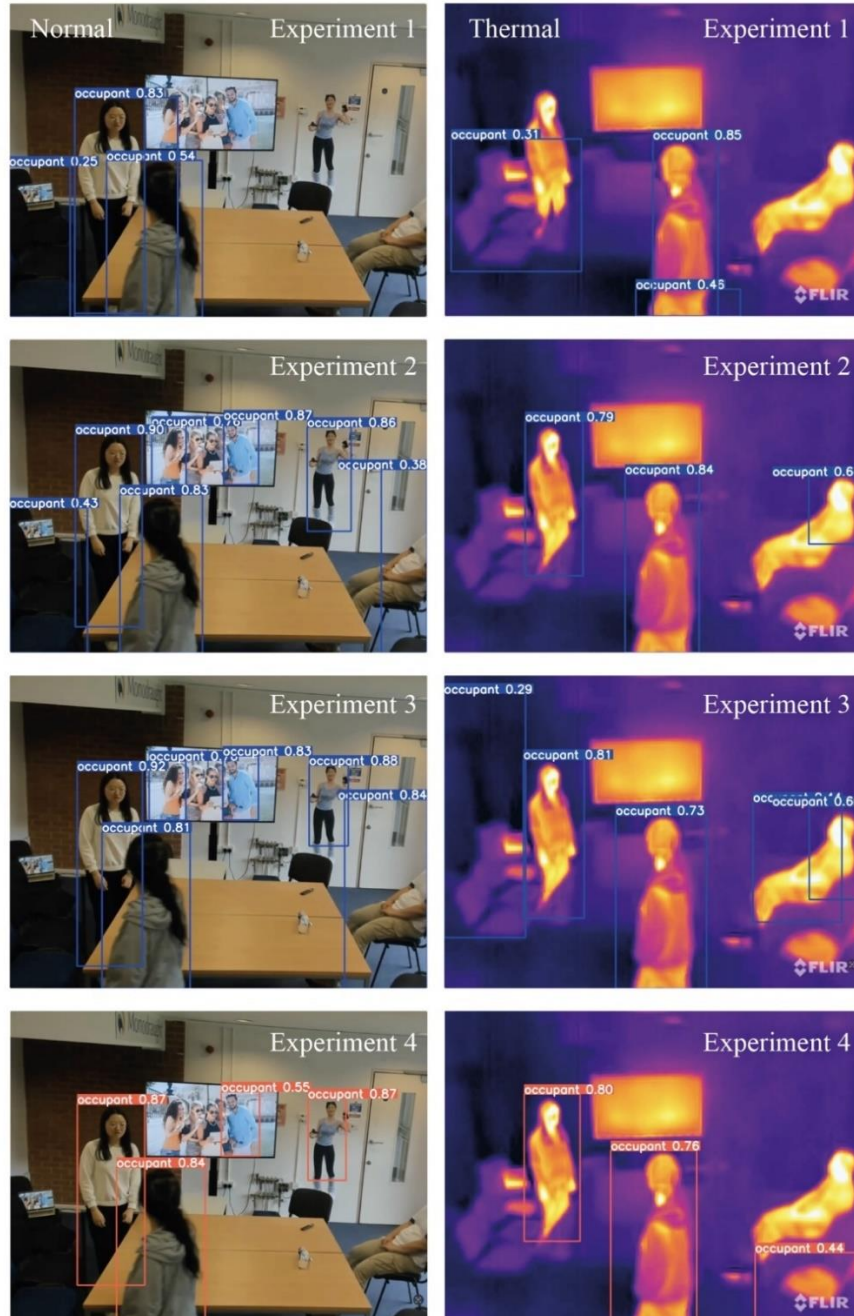


Figure 5-8 The Cross-Video Experiments video inference results in comparison for the standard and thermal models in the entering room stage.

In Experiment 1, as illustrated in the example screenshots in

Figure 5-8, a person seated on the right was not detected by either the standard or the thermal camera. With more training images in Experiment 2 and Experiment 3, both cameras were able to identify all occupants in the room. Despite this progress, the standard camera encountered challenges in scenarios where portrait pictures were placed on the wall. These images were occasionally misidentified as occupants, reducing the standard camera's detection accuracy. The thermal camera, in contrast, avoided such misdetections due to its design, which focuses solely on objects' heat. This advantage allowed it to maintain higher accuracy in scenarios involving visual distractions, such as portrait images. However, an unexpected issue arose with the thermal camera in Experiment 1 as shown in Figure 5-9, where it occasionally misidentified the heat signature of a monitor as an occupant. Although this error occurred in only a few frames, it highlighted a potential limitation of the thermal dataset when dealing with non-occupant heat sources. This issue was effectively addressed in Experiments 2 and 3 by expanding the training dataset to include more diverse scenarios, thereby improving the model's ability to distinguish between occupants and objects emitting heat.

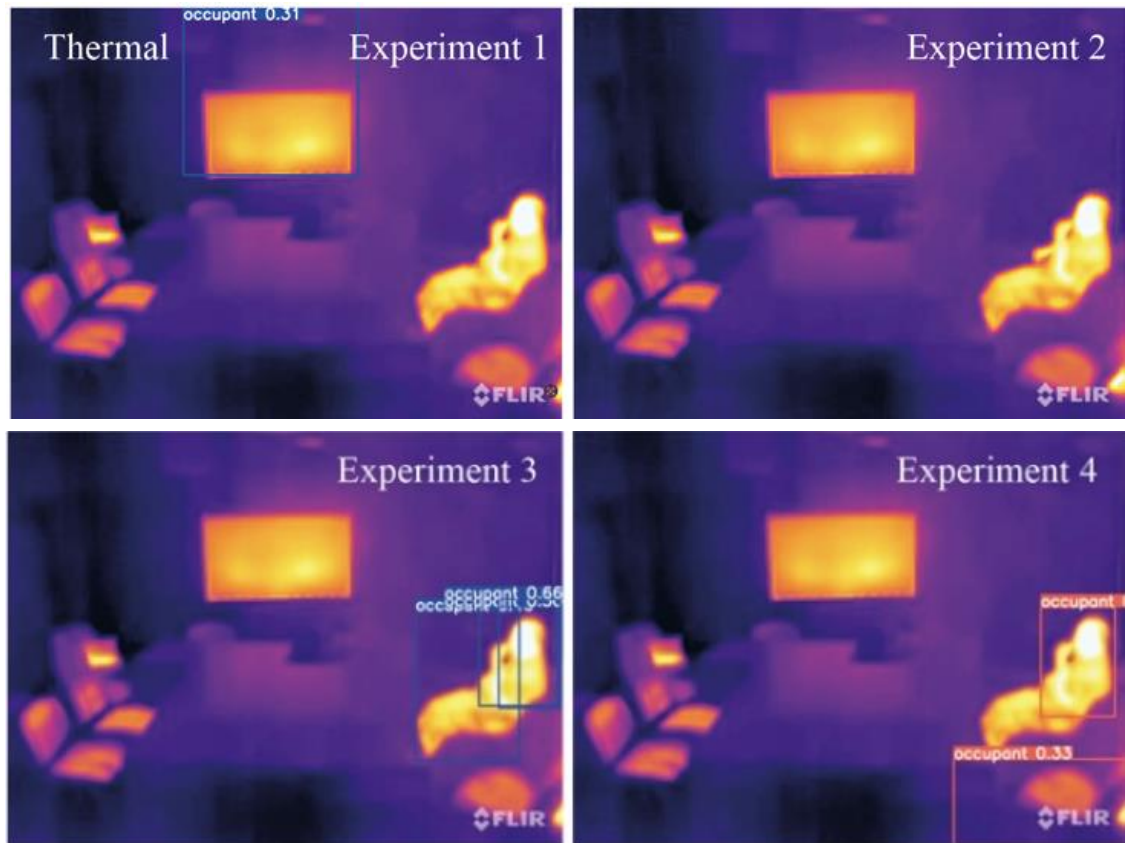


Figure 5-9 The detection results of thermal models at 01:14 of the inference video in Experiment 1 to 4.

In the second stage, where occupants were seated around a table for discussion, the detection accuracy in Experiment 1 was low, with both cameras achieving less than 50% accuracy. In Experiment 2, the standard camera demonstrated better performance than the thermal camera, as the thermal camera's accuracy continued to decline during this stage. Interestingly, the thermal camera's accuracy only improved toward the end of this stage when occupants began moving out of the room, aligning more closely with the ground truth. By Experiment 3, both cameras performed better and accurately detected occupancy numbers during this stage.

Example images of the detection results for the second stage are shown in Figure 5-10. In Experiment 1, both the standard camera and the thermal camera struggled to detect all occupants, particularly with overlapping occupants, which were detected as a single

entity. Additionally, neither camera successfully detected the person sitting next to the monitor, highlighting a shared limitation in the initial experimental setup. As a complex scenario, the images could be further analysed by dividing them into regions with overlapping and non-overlapping occupants. In Experiments 2 to 4, the non-overlapping occupants, such as the person seated next to the monitor, were successfully detected by both cameras. However, the regions with overlapping occupants remained challenging. A closer investigation of these areas revealed that occupants sitting on the right-hand side of the table were more spatially separated compared to those on the left-hand side. This separation contributed to improved detection accuracy for the right-hand side occupants in later experiments, as the increased diversity and complexity in the training datasets enabled the models to better handle such scenarios.

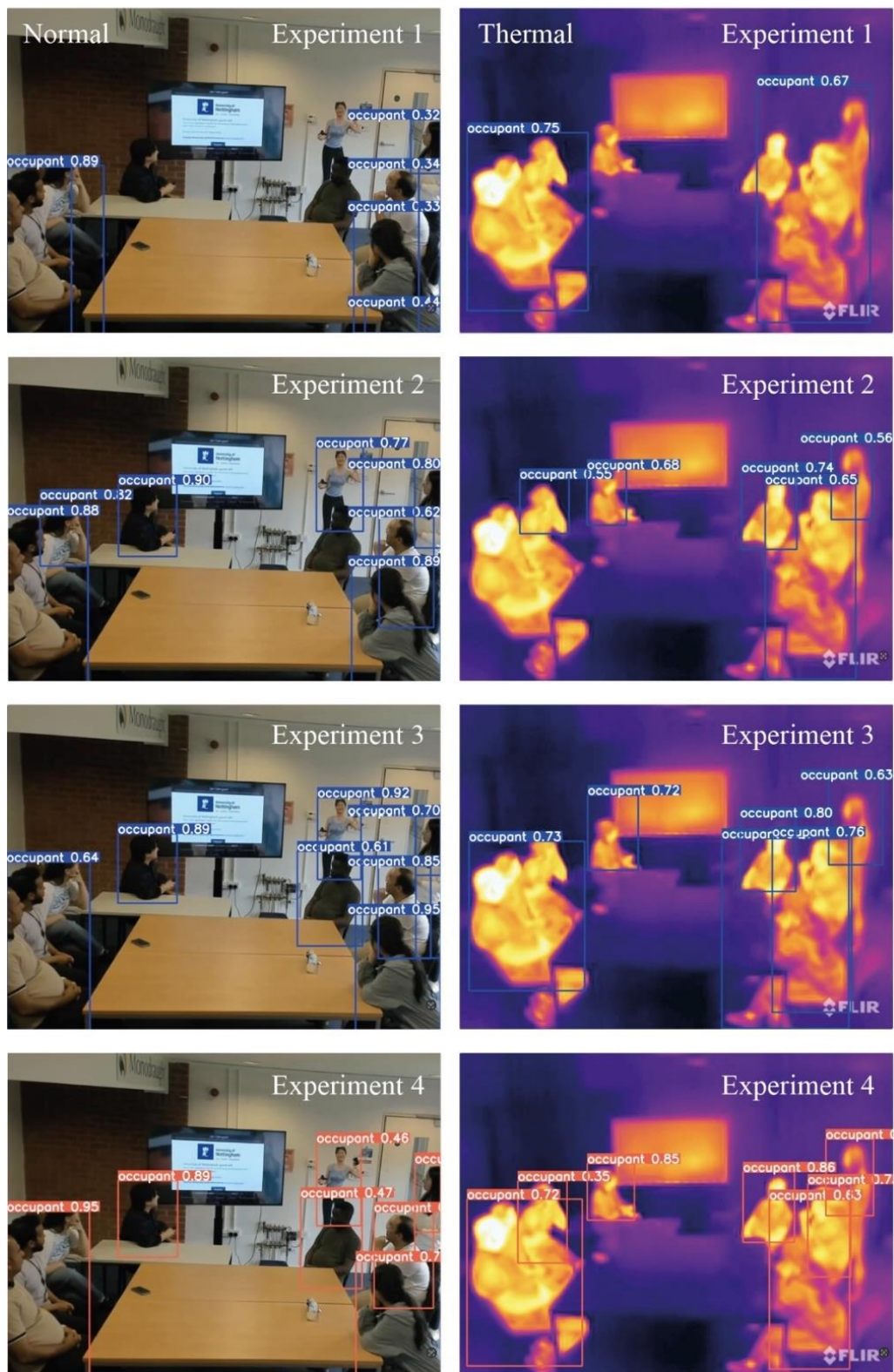


Figure 5-10 The Cross-Video experiments video inference results in comparison for standard and thermal models in the discussion (sitting) stage.

Some occupants went back to the room before they left in the third stage as shown in Figure 5-11 which made a temporary rise in the occupancy count for approximately one minute before declining once more. In Cross-Video Experiment 1, both the standard and thermal cameras demonstrated low accuracy during this stage. While the standard camera was able to capture the temporary increase in occupant numbers, its overall performance remained limited. Cross-Video Experiments 2 and 3 showed some improvements, particularly in Experiment 3, where the detection accuracy was better aligned with the ground truth, even as occupants moved back into the room and left again. The improved datasets used in Cross-Video Experiments 2 and 3 contributed to greater detection accuracy compared to Experiment 1. For example, in Cross-Video Experiment 3, the model was capable of detecting occupants positioned from the front of the view. In this stage, YOLOv8 outperformed YOLOv10, particularly in thermal camera applications. YOLOv10 displayed a tendency to misidentify heat-emitting objects as occupants, which reduced its accuracy. Additionally, there were instances where YOLOv10 failed to detect occupants standing in the background of the scene in a few frames. These limitations underline the strengths of YOLOv8, which provided more reliable detection of both near and far occupants, even under challenging conditions.

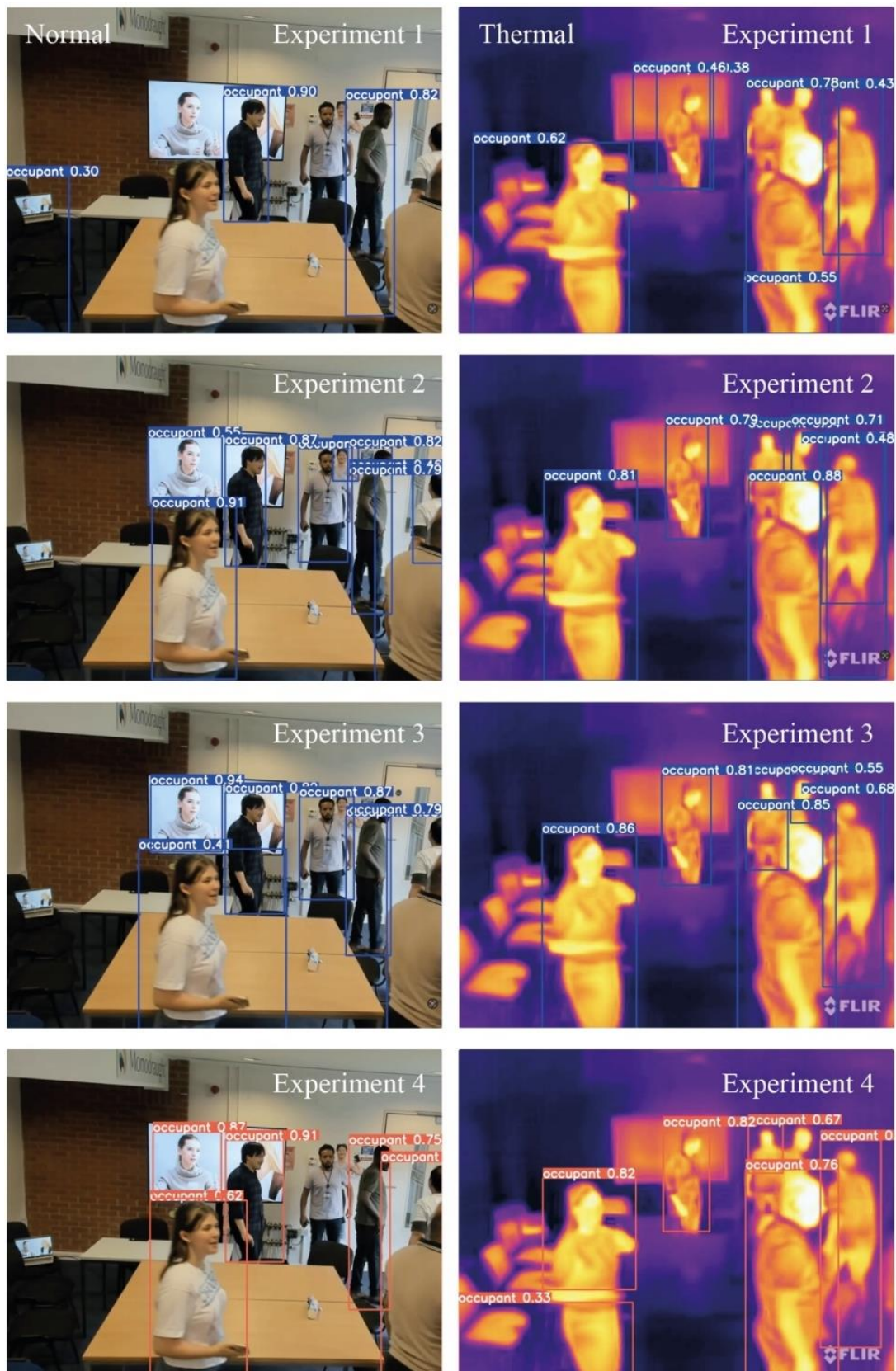


Figure 5-11 The Cross-Video experiments video inference results in comparison for standard and thermal models in the leaving room stage.

In the final stage, during which occupants were standing and engaging in discussion as in Figure 5-12, the overall accuracy of both the standard and thermal cameras was similar in Experiments 1 and 3. In Experiment 1, the standard camera tends to over-detect the number of occupants. Conversely, the thermal camera under-detected the number of occupants, likely due to challenges in distinguishing closely grouped individuals. By Experiment 3, both cameras over-detect, with the standard camera continuing to overestimate occupant numbers while the thermal camera showed slight inaccuracies due to overlapping heat signatures. Interestingly, in Experiment 2, the thermal camera outperformed the standard camera, achieving better accuracy with the ground truth during this stage. The standard camera, on the other hand, over-detected occupant numbers, reporting a count higher than the actual number. This discrepancy may be attributed to the presence of additional visual distractions, such as overlapping individuals or environmental factors, which impacted the standard camera's detection accuracy.

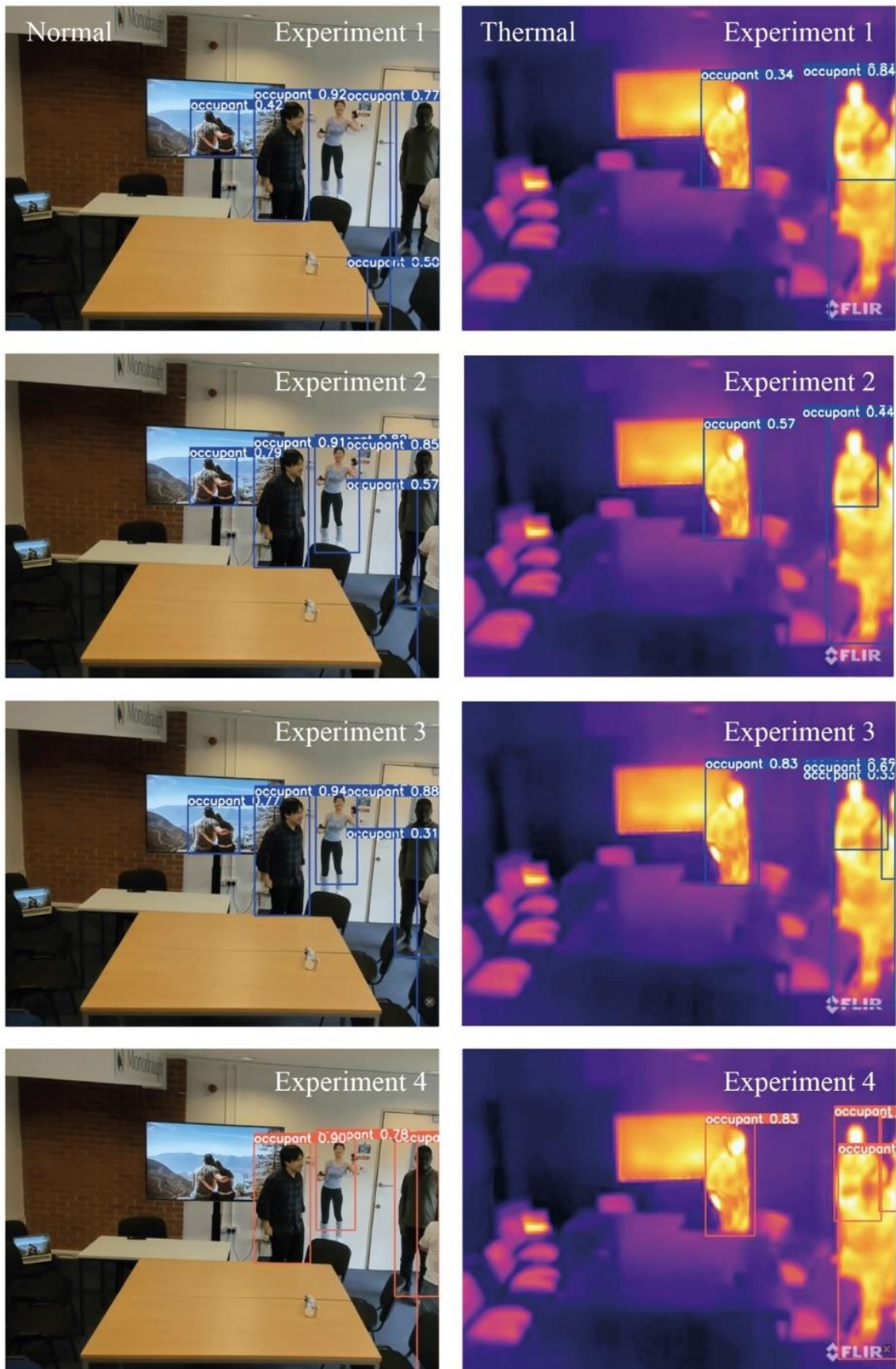


Figure 5-12 The Cross-Video experiments video inference results in comparison for standard and thermal models in the discussion (standing) stage.

5.4 Discussion

5.4.1 Metrics evaluation

Mean Average Precision (mAP) is a widely used metric in the field of object detection, serving as a standard for evaluating model performance during the training process. It provides a comprehensive measure by combining precision and recall across a range of intersection-over-union (IoU) thresholds, typically from 0.5 to 0.95 at 0.05 intervals. Precision refers to the proportion of correctly identified objects (True Positives) relative to all detected objects, while recall measures the proportion of correctly identified objects relative to the total number of ground truth objects. The mAP metric is useful for assessing model performance during training, as it evaluates both the localization and classification capabilities of the model under controlled conditions. A higher mAP score indicates that the model is capable of detecting objects with bounding boxes and correct labels, making it a reliable indicator of model effectiveness within provided datasets.

The mAP metric is particularly useful for assessing model performance during training and validation, as it evaluates both the localization and classification capabilities of the model under controlled conditions. A higher mAP score indicates that the model is capable of consistently detecting objects with precise bounding boxes and correct labels, making it a reliable indicator of model effectiveness within the scope of the provided datasets. However, while mAP is an essential tool during the training process, it has limitations when applied to unseen inference videos that more closely resemble real-world conditions. Training and validation datasets are often curated and structured, lacking the variability and complexity of practical environments. As a result, mAP may not fully capture the model's generalization ability or its robustness in dynamic scenarios.

To address this gap, manually counted accuracy calculated as the ratio of detected occupant numbers to the ground truth numbers within each 10-second interval of the inference video is introduced as a metric for evaluating the model's performance on inference videos. Unlike mAP, which focuses on precision at specific IoU thresholds, manual counted accuracy directly compares the number of detected occupants to the ground truth count in defined time intervals. This approach provides a more realistic measure of the model's effectiveness in unseen environments, where factors such as overlapping occupants, environmental distractions, and varying camera angles can influence detection accuracy. As in Figure 5-13a, by comparing mAP with manually counted accuracy, the gap between training and real-world inference performance can be quantified, offering insights into the model's adaptability and areas for improvement.

One challenge that affects both mAP and manual counted accuracy is the overlapping of occupants, which complicates both the labelling and detection processes. In scenarios where the body parts of one occupant are occluded by others, labelled occupants in the training set may only be partially visible in the testing set. This issue is in both standard and thermal cameras, as occupant overlap reduces detection accuracy. Furthermore, in thermal images, occupants may not have distinct heat signatures, making them harder to recognize and detect accurately. This limitation often results in labels that do not fully represent the actual occupants, introducing additional inconsistencies into the training and evaluation processes. Addressing these challenges requires larger dataset preparation and model enhancements that can account for complex scenarios involving overlapping and partially visible occupants.

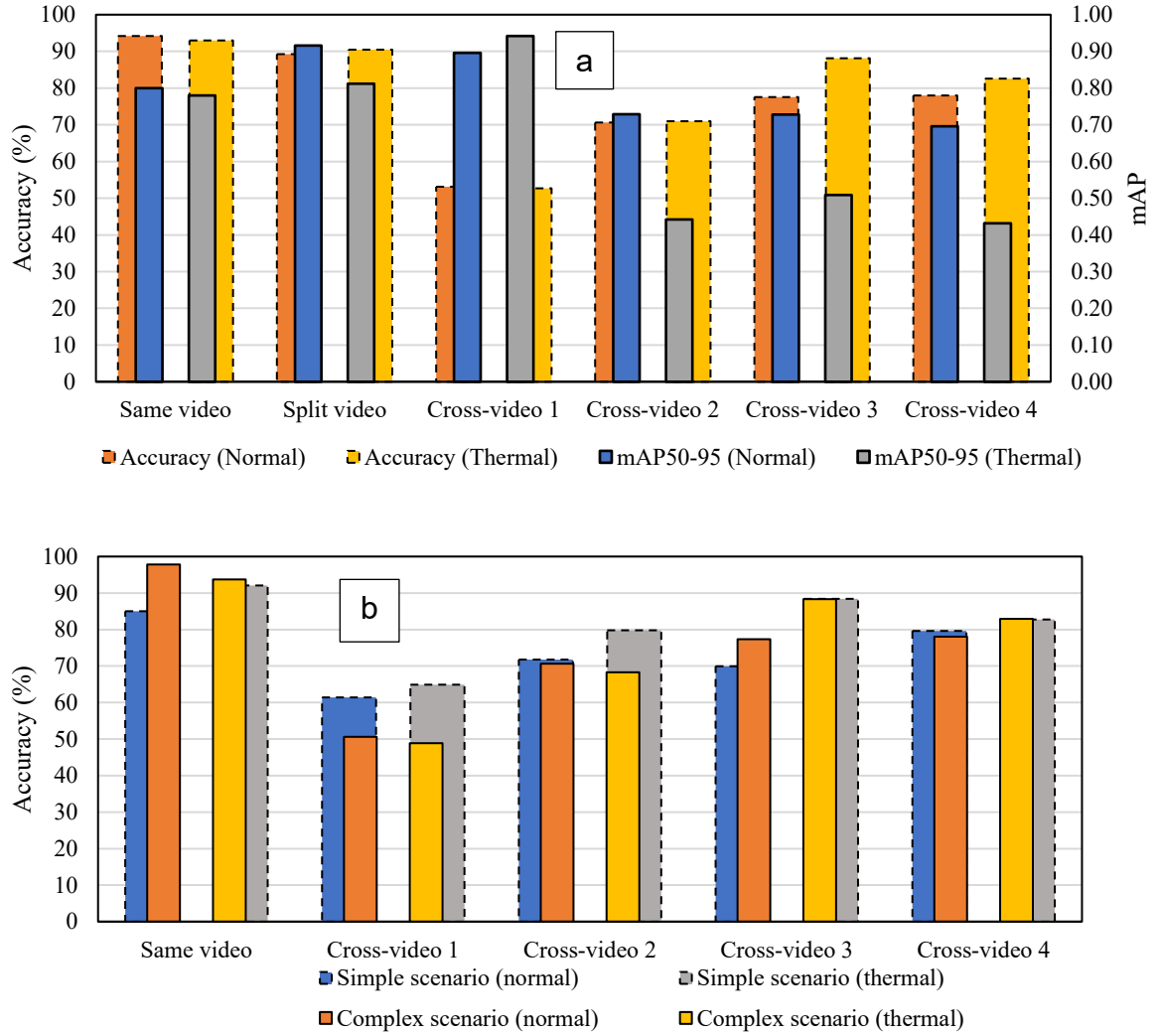


Figure 5-13(a) The comparison for mAP and manual counted accuracy in all experiments. (b) The accuracy comparison of simple and complex scenarios in different experiments.

In the same-video experiment, both cameras achieved approximately 94% accuracy and a mean average precision (mAP) of 0.8. This indicates that the highest performance of both cameras could be similar, and it could be achieved with a large training dataset. Manually counted accuracy tends to be higher than mAP because manual counting can be considered as an idealized mAP@100 metric, which is unattainable for a machine learning model but achievable in a manual process.

In the split-video experiment, accuracy dropped slightly for both cameras compared to the same-video experiment. However, the mAP increased due to the lower variation in scenarios, which included only two stages compared to four stages in the same-video experiment. In Experiment 1, mAP was significantly higher than manual counted accuracy, highlighting a substantial gap between model performance during training and its application in real-world prediction tasks. This discrepancy arose because both the training and testing datasets in Experiment 1 represented simple scenarios, whereas the inference video featured a mix of simple and complex scenarios. By contrast, in the split-video experiment, mAP closely aligned with accuracy. This was because the training dataset included images from stage one and the first half of stage two, encompassing both simple and complex scenarios, leading to improved generalization.

In the cross-video experiments, the gap between mAP and accuracy was smaller for the standard camera compared to the thermal camera, suggesting that the model trained with standard images exhibited better performance in terms of extracting the precise regions of occupants. Although labelling thermal images can be challenging due to difficulties in recognizing occupant outlines, this issue appeared to have a limited impact on mAP. Instead, the performance gap likely reflects the inherent limitations of the YOLOv8 model's ability when trained on thermal images. The difference in mAP between cross-video experiments 2 to 4 and the same-video experiment was larger for thermal camera detection, indicating that the thermal camera struggled to detect the exact location of occupants across diverse scenarios. Despite this, the thermal camera's ability to count occupants was unaffected, with manually counted accuracies reaching as high as 88% in cross-video experiment 3. This suggests that while the thermal camera may face challenges in precise localization, it remains effective for occupancy counting.

When considering occupancy profile generation, accuracy is arguably more critical than mAP. The number of occupants within a given interval directly influences the demand for heating or cooling, making accurate occupant counts more relevant for building

energy management systems. Additionally, the thermal camera offers the advantage of better privacy protection, as detailed facial features are not discernible in thermal images. While the standard camera demonstrated the potential to achieve slightly higher mAP and accuracy, both cameras delivered comparable performance in generating occupancy profiles, highlighting their utility for real-world applications with varying priorities.

5.4.2 Detection performance comparison in simple and complex scenarios

Simple scenarios which have low occupant density and not much overlap, represent controlled environments where detection is relatively straightforward. In contrast, complex scenarios, such as crowded scenes with occupant overlaps, introduce challenges that test the limits of the model's ability to detect and distinguish individuals accurately. The detailed comparison is shown in Figure 5-13b. By analysing performance in these contexts, the strengths and weaknesses of both the standard and thermal cameras can be identified, as well as the effectiveness of the training datasets in preparing the models for real-world applications.

In this study, Cross-Video Experiment 3 achieved the highest accuracy for both cameras. For the standard camera, errors caused by misdetections tended to have a greater impact in simpler scenarios due to the smaller number of total occupants. However, this trend was only consistently observed in Cross-Video Experiment 3. The lower accuracy observed in crowded scenarios during Cross-Video Experiments 1 and 2 suggests that the datasets used in these experiments lacked sufficient diversity and complexity. Additionally, YOLOv10 demonstrated better generalization capabilities with the standard camera, achieving approximately 80% accuracy in simple scenarios compared to YOLOv8's 70%, reflecting the benefits of improved model architecture for certain conditions. A similar pattern was observed for the thermal camera. In Cross-Video Experiments 1 and 2, the accuracy in crowded scenarios was lower than in simple scenarios and improved with the increasing dataset. In Cross-Video Experiments 3 and

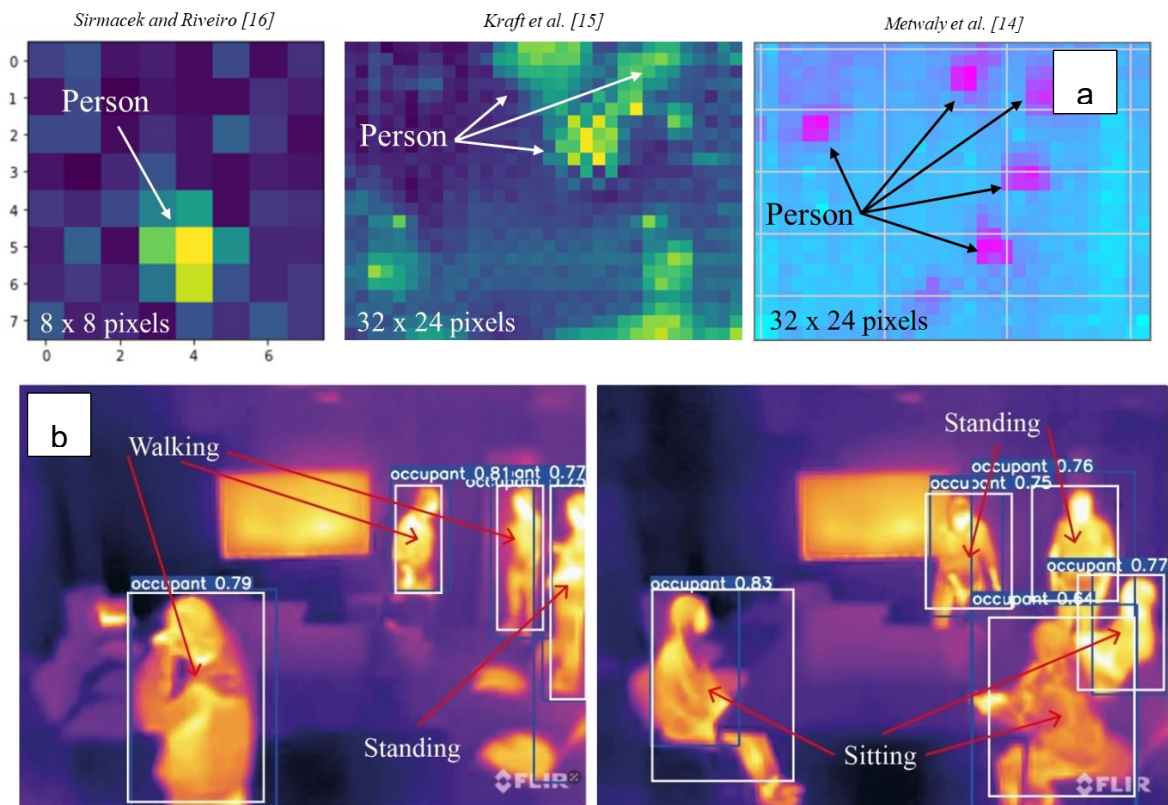
4, the detection accuracies in simple and crowded scenarios are the same, indicating the model's ability to detect occupants in complex scenarios with Dataset 3. This highlights the importance of dataset diversity and coverage for enhancing model performance in complex scenarios.

Despite the improvements, a gap remains between the highest accuracy achieved in cross-video experiments and same-video experiments. Considering the limitations in dataset size and variability, the difference of approximately 6% is acceptable. This small gap suggests that, with sufficient dataset preparation, both standard and thermal cameras can achieve near-optimal performance in challenging real-world scenarios. The results also underscore the importance of datasets to include diverse scenarios, particularly for models like YOLOv8, which require comprehensive training data to generalise effectively. While YOLOv10 demonstrated slightly better in certain cases, the choice between these models should consider specific application requirements, including accuracy, dataset availability, and computational efficiency.

5.4.3 Additional findings in vision-based occupancy prediction

Most existing vision-based occupancy prediction studies use standard cameras as sensors, with limited research exploring the use of thermal cameras. In this study, we addressed this gap by testing and comparing the performance of standard and thermal cameras for occupancy prediction. The comparison provided valuable insights into the advantages and limitations of each sensor in various scenarios. Additionally, we uncovered some novel findings for vision-based occupancy prediction, demonstrating its potential in privacy-sensitive applications and dynamic or low-light environments. These findings not only highlight the feasibility of thermal cameras for occupancy detection but also emphasize the importance of further exploring their capabilities to expand the scope of vision-based methods in real-world applications.

As mentioned earlier, most previous studies have relied on low-resolution thermal cameras, which are limited in their ability to capture occupants with detail (Figure 5-14a). These cameras often render occupants as indistinct "blobs," making it challenging to detect finer features or analyse specific behaviours. In contrast, this study employed a higher resolution but cost-effective thermal camera, the FLIR ONE Pro, which provides detail and improves the accuracy of occupant detection. As shown in Figure 5-14a, the deep learning model trained on data from the camera could detect occupants when they are in different behaviours, including walking, standing, and sitting. This highlights the potential of leveraging higher-resolution thermal cameras for more advanced tasks beyond simple occupancy counting, addressing the limitations of previous studies while maintaining affordability and practicality for real-world applications.



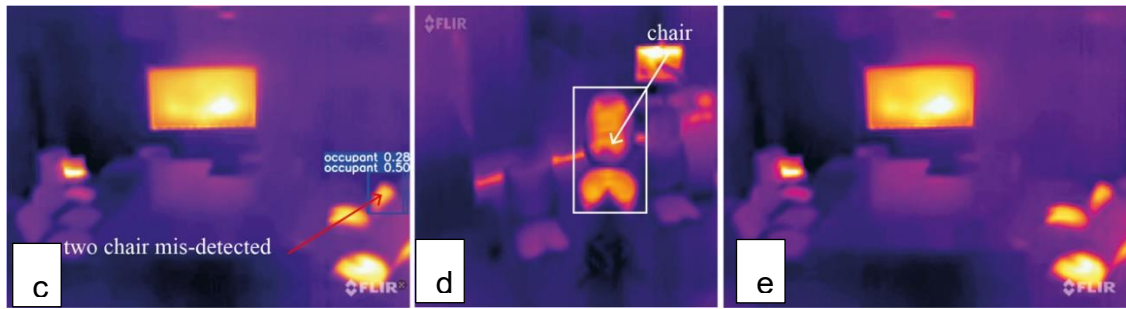


Figure 5-14 (a) Examples of occupancy detection in previous studies, and (b) example pictures of deep learning model detecting occupants in the present study. The thermal imprints left on chairs: (c) the deep learning model mis-detected it as an occupant in Cross-Video Experiment 3, (d) additional images were added to the dataset with thermal imprints left on chairs, and (e) the model correctly ignored the thermal imprints left on chairs.

In this study, we observed thermal imprints left on chairs in the thermal images, where the residual heat signature persisted for a period after the occupant had left. This phenomenon posed a challenge for accurate detection, as the model often misclassified these heat signatures as occupants. For example, in Experiment 3, as shown in Figure 5-14b, the model incorrectly detected the residual heat in a chair as a person. To address this issue, additional images of chairs with thermal imprints were added to the training dataset. Specifically, 211 such images were included, as illustrated in Figure 5-14c. This adjustment allowed the model to learn to distinguish residual heat patterns from actual occupants. The results, as shown in Figure 5-14d, indicated that the model no longer mis-detected the residual heat signature as an occupant.

However, while this improvement resolved the specific issue of residual heat misdetection, it came at the cost of a decrease in overall accuracy. The accuracy dropped from 88% to 76%, indicating that the addition of these images may have impacted the model's ability to generalise across other scenarios. This suggests that while targeted additions to the training dataset can help address specific challenges, they may also introduce trade-offs in overall performance. Future work should explore ways to balance

dataset specialization with generalization, such as fine-tuning models for specific tasks or employing advanced techniques to reduce the impact of dataset biases.



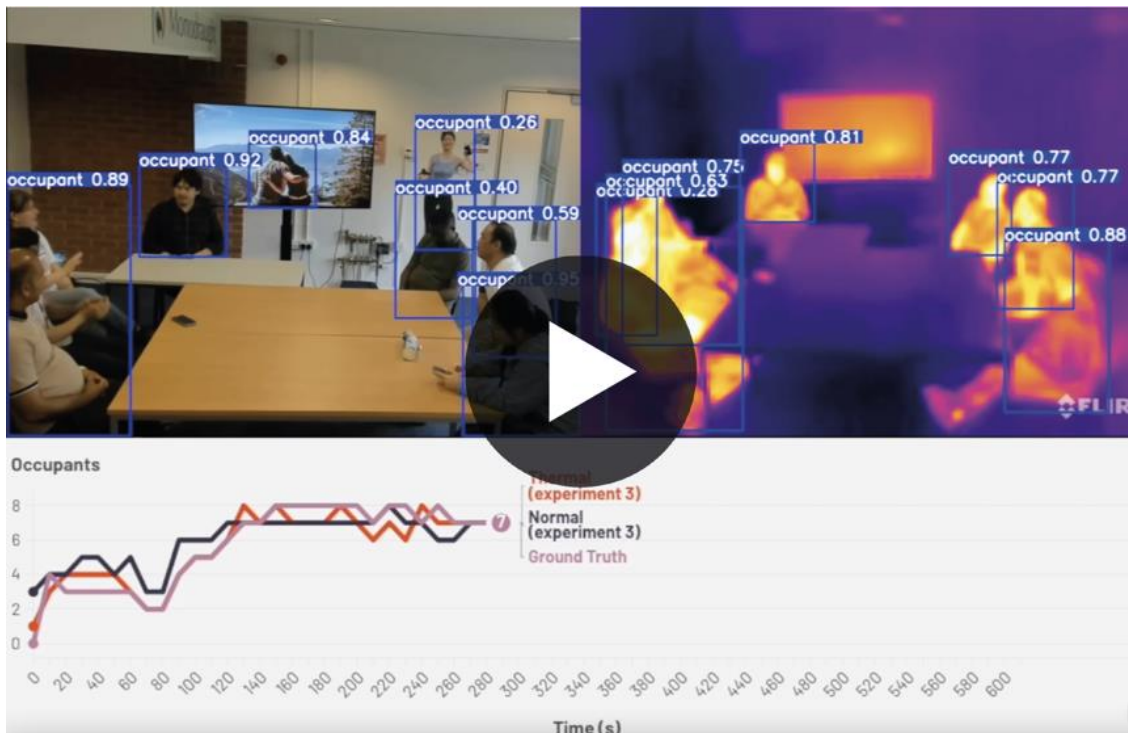
Figure 5-15 The detection of overlapping people. The dataset of Experiment 2 added more images with people in different environment and Experiment 3 added crowded scenarios specifically.

Overlapping occupants is always a challenge for vision-based occupancy detection because they create ambiguity in distinguishing individuals within the same region of

an image. For thermal imaging, the heat signatures of closely positioned occupants often merge, resulting in blended patterns that obscure individual outlines. Similarly, in standard camera footage, overlapping can cause visual features such as limbs or body contours to overlap, making it difficult for detection models to separate one person from another. These challenges affect both the labelling process during dataset creation and the model's ability to generalise to real-world scenarios, especially in crowded environments or dynamic settings where occupants move or interact closely.

As illustrated in Figure 5-15, in Cross-Video Experiment 1, both standard and thermal cameras struggled with overlapping occupants, failing to detect individuals separately in complex scenarios. However, as additional images containing overlapping occupants were incorporated into the training dataset for Experiment 3, the thermal camera demonstrated improvement, as shown in Video 5-1. It was able to detect overlapping individuals more accurately, with bounding boxes correctly identifying separate occupants. In contrast, while the standard camera performed relatively well in Experiment 1, its performance did not improve much in the following experiments, suggesting that it may have already reached its detection capability for overlapping scenarios, possibly due to the reliance on visual features that become ambiguous in such conditions. This highlights the limitations of standard cameras when dealing with visual overlap, as they rely on distinct body contours and visual details that are often obscured in crowded or overlapping situations.

The improvement observed in the thermal camera underscores the importance of dataset diversity, particularly in addressing overlapping scenarios. By including more representative examples in the training data, the thermal camera was better equipped to generalise and separate overlapping heat signatures. These findings suggest that while standard cameras are effective under control or simpler conditions, thermal cameras, with adequate training, can offer another option for detecting occupants in complex and crowded environments.



Video 5-1 The comparison of ground truth, standard and thermal deep learning models in Cross-Video Experiment 3. *The playable video is available at <https://youtu.be/b6MnmFJNZ2E>*

5.5 Summary

This study explored and compared the performance of standard and thermal cameras for vision-based occupancy detection, both cameras can reach around 70% accuracy with sufficient dataset preparation employing YOLOv8 and around 80% accuracy with YOLOv10 models. Through Same-Video, Split-Video, and Cross-Video experiments, the findings demonstrated the strengths and limitations of both camera types in occupancy prediction, highlighting their ability to detect people in various positions, such as walking, sitting, and standing, across both simple and complex scenarios.

The results revealed that both cameras achieved approximately 94% accuracy and a mean average precision (mAP) of 0.8 in controlled settings, which demonstrated their maximum potential, as in the Same-Video experiment, where the training data were duplicated with the validation video. However, in more challenging scenarios, such as

the Cross-Video experiments, thermal cameras faced greater difficulties in distinguishing fine details, such as overlapping occupants, due to the blending of heat signatures. Despite these limitations, thermal cameras exhibited competitive performance, especially in privacy-sensitive applications, by effectively avoiding visual distractions like portrait images or facial features present in standard camera datasets.

Key insights from this study include the critical role of dataset diversity and selection in improving model performance, particularly in scenarios involving overlapping occupants or residual heat imprints. The results demonstrated that increasing the size and complexity of training datasets significantly improved the model's ability to generalise and adapt to real-world conditions. For example, the inclusion of diverse overlapping scenarios in Experiment 3 enabled the YOLO model to accurately detect individual occupants, even in crowded environments, highlighting the importance of comprehensive training data in addressing complex challenges.

Additionally, this study underscores the potential of cost-effective, higher-resolution thermal cameras, such as the FLIR ONE Pro, for enhancing vision-based occupancy prediction. With their privacy-preserving capabilities and improved detection of occupant behaviours, thermal cameras can expand the scope of applications in smart building management, energy efficiency, and occupant monitoring. However, challenges remain, particularly in achieving consistent accuracy across diverse and dynamic conditions, such as scenarios with overlapping occupants or residual heat imprints.

In conclusion, while both standard and thermal cameras demonstrated comparable performance under optimal conditions, thermal cameras offer unique advantages in privacy-sensitive and low-light environments. This study contributes to bridging the research gap in thermal-based occupancy detection by highlighting the strengths and limitations of these cameras and providing a foundation for future research, for example, the next chapter explored occupancy thermal comfort prediction with the vision-based

deep learning method by using the comfort level in thermal image as detection objectives. To further enhance the performance and generalization of thermal imaging models, future efforts should focus on developing more diverse training datasets, advanced detection techniques, and methods to address challenges such as residual heat and overlapping occupants. By addressing these gaps, vision-based thermal occupancy detection can become a more reliable tool for real-world applications including more specific objectives.

6. DYNAMIC THERMAL COMFORT PREDICTION WITH THERMOGRAPHIC IMAGING

6.1 Introduction

In the last chapter, the deep learning model with thermal images showed good performance with appropriate datasets which provides potential for more specific applications. This chapter aims to develop a real-time, vision-based deep learning model for detecting thermal comfort levels in building environments using thermal imaging technology. The research focuses on utilising the YOLOv8 algorithm, a state-of-the-art single-shot object detection system renowned for its efficiency in real-time applications (Jocher, 2023). YOLO-based algorithms have demonstrated promising results in indoor settings, such as a 2022 study on occupancy counting with YOLOv4, which achieved an accuracy of 96–99% (Lee et al., 2022). Similarly, another study using YOLOv5 for occupancy counting in two office environments also reported high accuracy (Choi et al., 2021). Building on this foundation, this research seeks to address existing limitations by developing a method that automates the analysis of thermal images, eliminating the need for manual input during feature extraction. Unlike traditional approaches that rely on manually defined ROIs, this method employs a vision-based deep learning model to automatically detect and process thermal patterns directly from raw data, offering a fully automated and adaptable solution.

While traditional models such as the Predicted Mean Vote (PMV) and the adaptive model have formed the basis for thermal comfort assessment in buildings, their generalised nature often fails to reflect the diverse and dynamic responses of individual occupants. These models typically rely on steady-state environmental and metabolic assumptions, which limit their sensitivity to real-time changes in personal comfort. In response to these limitations, recent research has begun to explore dynamic thermal

comfort modelling, with real-time data and individual responses. This chapter developed a personalised thermal comfort prediction model using thermographic imaging and deep learning which aims to capture individual variability in comfort levels more accurately, providing a foundation for future occupant-centred and adaptive building control systems. Although this method is exploratory and still in development, it represents a shift towards more intelligent and responsive comfort modelling that can adapt to occupant needs in real time.

Field experiments were conducted in a real-world office environment to collect thermal image data from multiple subjects, along with Thermal Sensation Votes (TSVs) recorded during controlled temperature variations. These thermal images were labelled with their corresponding TSVs, forming a dataset used to train a deep learning-based thermal comfort prediction model. The primary goal of this study is to assess the feasibility of using vision-based deep learning methods to predict individual thermal comfort in real-time. To evaluate the feasibility of the model focusing on generalisability and accuracy, cross-validation techniques were also employed, including intra-subject and cross-subject validation. Intra-subject validation tested the model on data from the same individuals included in the training, while cross-subject validation evaluated its ability to predict thermal comfort levels for unseen occupants.

The model's accuracy and performance will be compared to the traditional PMV approach, which typically relies on environmental measurements such as temperature, humidity, and air velocity. This comparison aims to provide initial insights into the potential of the vision-based method as a flexible and cost-effective alternative to traditional thermal comfort detection techniques.

A novel method for detecting thermal comfort is introduced using a vision-based deep learning approach, addressing significant limitations in existing thermal comfort prediction techniques. The experimental setup involves 14 field experiments conducted

in an office environment, during which data were collected from subjects, including thermal images, TSVs, and environmental information. These thermal images, paired with their corresponding TSVs, form a foundational dataset for training the model, which predicts real-time occupancy thermal comfort levels and compares them with PMV values calculated from environmental data.

A key contribution lies in the development of a fully automated vision-based method that eliminates the need for environmental sensors or manual feature extraction. Unlike traditional approaches, which require manually defining regions of interest (ROIs), the proposed model leverages a deep learning framework to process raw data from a low-cost thermal camera, automatically identifying relevant patterns and extracting features. By automating these processes, the method reduces the reliance on specialised knowledge, offering a scalable, efficient, and cost-effective solution for real-time thermal comfort prediction.

Additionally, this chapter advances the field by demonstrating a practical alternative to traditional PMV-based methods, which often depend on complex, multi-sensor setups. By relying solely on a low-cost thermal camera to capture the necessary data, the proposed approach simplifies the system, reduces costs, and enables broader application in real-world building environments. This chapter introduces a methodology that combines cutting-edge deep learning with accessibility and scalability, making real-time thermal comfort prediction feasible for diverse contexts.

6.2 Thermal comfort prediction method

The methodology involves evaluating the performance of a vision-based deep learning method for predicting thermal comfort levels in building environments and comparing its effectiveness to the traditional PMV model. Experiments were conducted with single occupants in an air-conditioned space during winter, beginning with the heating system

turned off. Each experiment lasted approximately two hours. The room gradually heated using an air conditioner (AC), causing the theoretical PMV to transition from cold to neutral to hot until the AC reached its heating limit. The AC was then switched to cooling mode, reversing the cycle from hot to neutral to cold.

Indoor environmental data and TSVs were collected every five minutes, and thermal video footage was recorded throughout each experiment. According to the ASHRAE Standard-55 (Standard, 1992), the TSV scale was based on a 7-point system: “-3” = cold, “-2” = cool, “-1” = slightly cool, “0” = neutral, “+1” = slightly warm, “+2” = warm, and “+3” = hot. However, the thermal states of “cold” and “hot” were not observed in most cases, resulting in datasets that typically included only four or five categories.

The thermal video recordings from the heating phase, encompassing cold-neutral-hot transitions, were segmented into frame-by-frame images. These images, paired with the corresponding TSVs, were labelled to create a dataset for training the deep learning model. The video data from the cooling phase were used to test the trained model using data from the same subject, representing the intra-subject test. The cross-subject test, which involves data from different individuals, is discussed later. The workflow for the intra-subject model is presented in Figure 6-1.

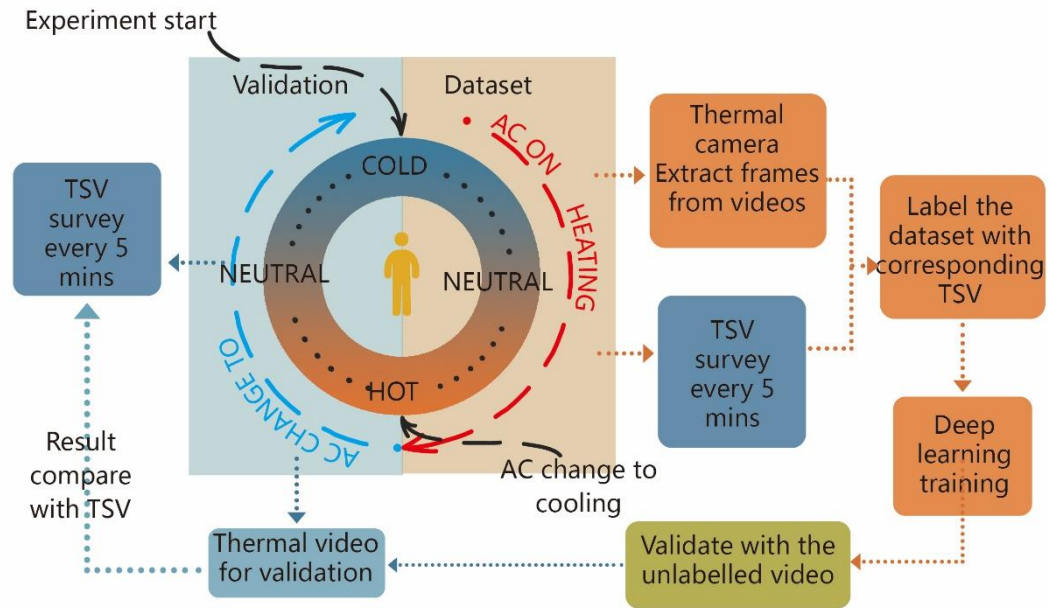


Figure 6-1 The workflow for the intra-subject thermal comfort detection model evaluation.

This study includes 14 separate experiments, each involving a different subject, to develop and validate a deep-learning model for thermal comfort detection. Each experiment is treated as an individual case, with the data split into two parts: one half is used for training the model, while the other half is used for validation. This intra-subject analysis evaluates the model's ability to predict thermal comfort for the same subject under controlled conditions.

In addition to intra-subject analysis, the study incorporates a cross-subject analysis. In this approach, data from multiple subjects are combined to train a generalised model, which is then tested on data from different, unseen subjects. This allows the preliminary evaluation of the model's capacity to predict thermal comfort levels across various individuals, providing insights into its adaptability for broader, real-world applications. By combining intra-subject and cross-subject approaches, the study investigates both subject-specific performance and the model's potential to generalise across diverse

occupants in building environments. In total, 212 TSVs and environmental data points were collected from 14 subjects to build and test the models. The detailed procedure for the vision-based deep learning thermal comfort prediction workflow is presented in Figure 6-2.

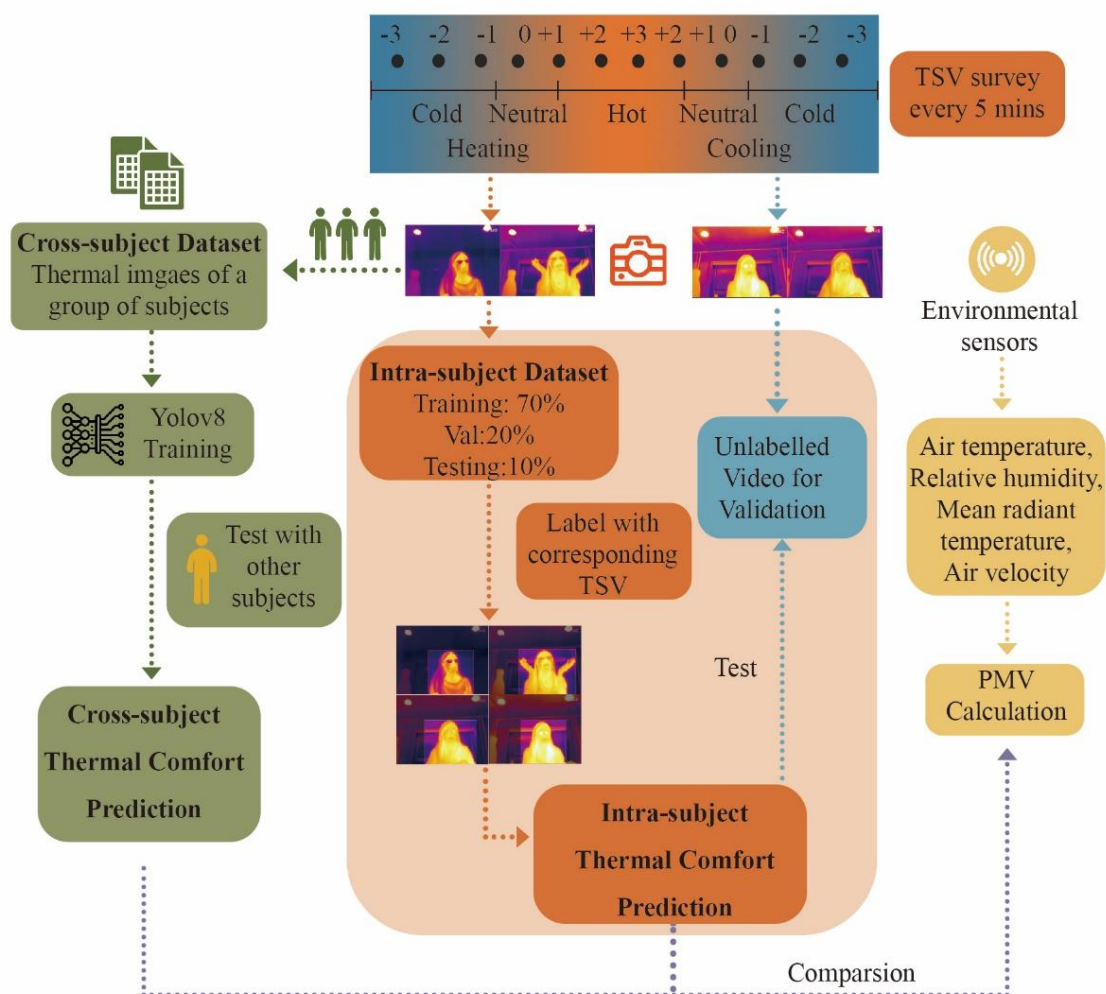


Figure 6-2 Detailed workflow for the proposed vision-based thermal comfort detection method.

6.2.1 Case study room setup

Experiments were conducted in an air-conditioned room within the Sustainable Research Building at the University of Nottingham, UK, which is located in a temperate

oceanic climate (Köppen: Cfb). The case study room, a meeting space accessible to staff and students in the Department of Architecture and Built Environment from 9:00 to 18:00 is illustrated in Figure 6-3. The experiments took place between November 2023 and February 2024, during typical winter weather in the UK. Outdoor temperatures ranged from -6°C to 16°C, with an average of 7°C, and humidity levels varied between 48% and 100%. The building is constructed to a BREEAM Excellent standard, with a U-value of 0.15 W/m²K for the roof and floor, 0.17 W/m²K for the walls, and 1.92 W/m²K for the windows.

The case study room measures 8.85 meters in length, 5.6 meters in width, and 2.45 meters in height. It features two windows on the east and west sides and one on the south. Further details of the room setup can be found in (Wei et al., 2022b).



Figure 6-3 (a) The Sustainable Research Building and (b) overview of the case study room.

A split wall-mounted air conditioner with both cooling and heating capabilities (Fujitsu; Cooling Capacity: 7.1 kW, Heating Capacity: 8.0 kW) was installed to regulate the room temperature. The air conditioner's temperature setpoint range is 18°C–30°C, with the actual room temperature during the experiments varying between 12°C and 30°C. During the heating period, the setpoint was maintained at 30°C, while it was adjusted to 18°C during the cooling period. It is important to note that the air conditioner was operated at a low speed, and care was taken to ensure that the air outlet did not blow directly toward the occupant, minimising any potential influence on thermal sensation or comfort levels.

6.2.2 Experimental setup and procedure

Throughout the experiment, three primary categories of data were gathered: environmental data (including observations on clothing and activity), thermal videos

and images, and Thermal Sensation Votes (TSVs). The environmental data were used exclusively to calculate the PMV for comparison with the proposed vision-based method using thermal images. A global thermometer (Tenmars TM-188) and a hot-wire anemometer (Model 440i, Testo Inc.) were positioned at a height of 1.1 m (seating height). The global thermometer measured air temperature (T_a), globe temperature (T_g), and relative humidity (RH) with a measuring frequency of 1-minute, averaging data every 5 minutes. The hot-wire anemometer measured indoor air speed with a measuring frequency of 1 second and provided manual averages every 5 minutes. The experimental setup is shown in Figure 6-4, and the detailed range, resolution, and accuracy of the instruments are listed in Table 6-1. All participants were instructed to wear one layer of clothing. Consequently, when environmental data were acquired, the corresponding PMV was calculated and recorded. The detailed PMV calculation method is described in Section 6.3.1.

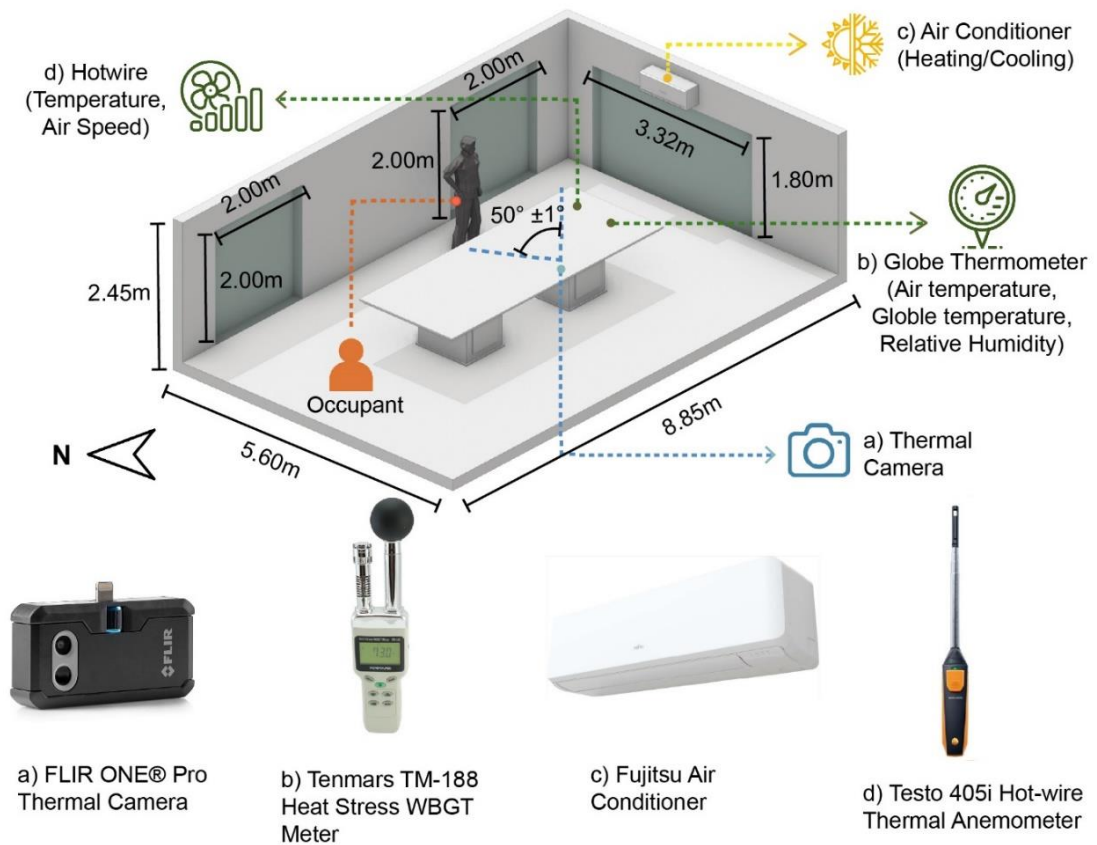


Figure 6-4 The thermal camera and environment sensors' location in the case study room.

Table 6-1 Sensors for collecting environmental data in the field experiment.

| Measurement parameters | Sensor | Range | Resolution | Accuracy |
|------------------------|---|-------------|------------|-------------------------|
| Air Temperature | Heat Stress WBGT Meter (Tenmars TM-188) | 0-50°C | 0.1°C | ±0.8 °C |
| Globe Temperature | | 0-80°C | 0.1°C | ±0.6 °C |
| Relative Humidity | | 1%-99% | 0.1% | ±3% |
| Air Speed | Air Flow Anemometer Testo 405i | 0 to 30 m/s | 0.01 m/s | ± (0.1 m/s + 5 % of mv) |

In each experiment, a thermal camera (FLIR ONE Pro) was positioned directly in front of the subject to capture thermal images of the subject's body. The camera was kept in a fixed location throughout all experiments to ensure consistency in data collection. Subjects were free to stand, sit, or walk, provided they remained within the camera's field of view. The FLIR ONE Pro was chosen for its availability and affordability

(costing approximately £300-400), which will be critical for the wider deployment of this technology in buildings. While the camera captures thermal images at a low resolution (160×120), this is advantageous for ensuring privacy while still capturing essential temperature variations and patterns. The camera operates within a spectral range of 8–14 μm and has a horizontal field of view of $50^\circ \pm 1^\circ$. Its accuracy is $\pm 3^\circ\text{C}$ or $\pm 5\%$, with an object temperature range of -20°C to 120°C . The emissivity of the thermal camera was set to 0.98, which is suitable for human skin (Ammer and Ring, 2019), and the reflection temperature was set to 22°C , the default setting of the FLIR ONE Pro.

The thermal camera has automated flat-field correction and is calibrated in the factory (Aryal and Becerik-Gerber, 2019). In this procedure, the shutter shuts off and the auto-calibration is carried out every three minutes using a uniform thermal scene. The infrared (IR) scale was set to 5°C to 40°C across all experiments, as temperature changes in this range were of interest. The thermal camera recorded videos of temperature changes and occupant movements. Portions of the video were converted into individual images to create datasets for training deep learning models. The remaining video data were reserved for validation purposes.

For this experiment, 14 healthy subjects were gathered, 8 of whom were female and 6 of whom were male. They were all international students from Asia, Europe, and Africa, and their ages ranged from 25 to 35. All subjects had no prior history of skin or cardiovascular conditions, and they had to abstain from alcohol, stay up late, take medication, and engage in strenuous activities 12 hours before the experiment. (Yao et al., 2008). During the experiment, the participants were dressed in long levis and trousers, of which the clothing insulation was around 1.0 clo and in a sedentary state (metabolic rate was around 1.0 met).

All participants were informed about the experiment and provided their consent to be recorded and complete the survey before the experiment. They were instructed to behave

as they normally would in their daily lives and were free to adopt any position within the camera's field of view. The air conditioner (AC) began heating the room at the start of the experiment and switched to cooling mode once the temperature reached the AC's upper limit. The questionnaire consisted of two parts. The first section introduced the research, explaining the concepts of PMV and TSV. The second section comprised the TSV survey, which the researcher reminded participants to complete every 5 minutes. Further details of the questionnaire are provided in the Appendix. The questionnaire was designed to focus solely on TSV, keeping the survey straightforward for participants. The TSV responses aligned directly with a seven-point scale ranging from cold (-3) to hot (+3), corresponding to the PMV index. To ensure the privacy and comfort of participants, no personal information or identifiable images were collected during the study.

6.2.3 Thermal camera calibration experiment

In this study, the FLIR ONE Pro thermal camera used for occupancy detection was factory-calibrated, ensuring baseline accuracy for temperature measurements. However, to further validate its performance and assess any potential deviations, an additional calibration experiment was conducted following the manufacturer-recommended method (Glavaš, 2024). This experiment aimed to compare the camera's temperature readings with a Pico Technology high-temperature Type K Thermocouple, which served as the reference instrument for precise temperature measurements. The setup, as shown in Figure 6-5, included three different thermal conditions: ice (cold reference), hot water (hot reference), and ambient air (moderate reference).

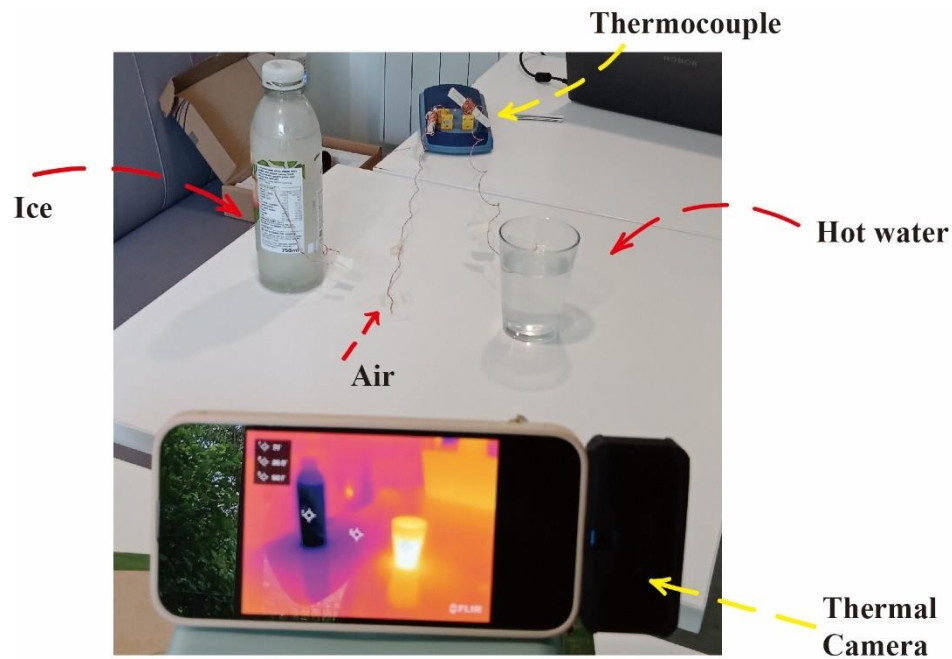


Figure 6-5 Experimental setup for thermal camera calibration using high-temperature type k thermocouples.

In this experiment, the thermal camera was positioned to capture all three objects while the thermocouple simultaneously recorded their actual temperatures. The ice was placed in a sealed bottle to maintain a stable low temperature, while the hot water served as the high-temperature reference. The thermocouple was in contact with the ice and the hot water to provide precise measurements, while the thermal camera relied on infrared radiation emitted by the surfaces of the objects. Both devices recorded temperature values every minute for a total duration of nine minutes, capturing any changes over time to assess consistency and potential drift in readings.

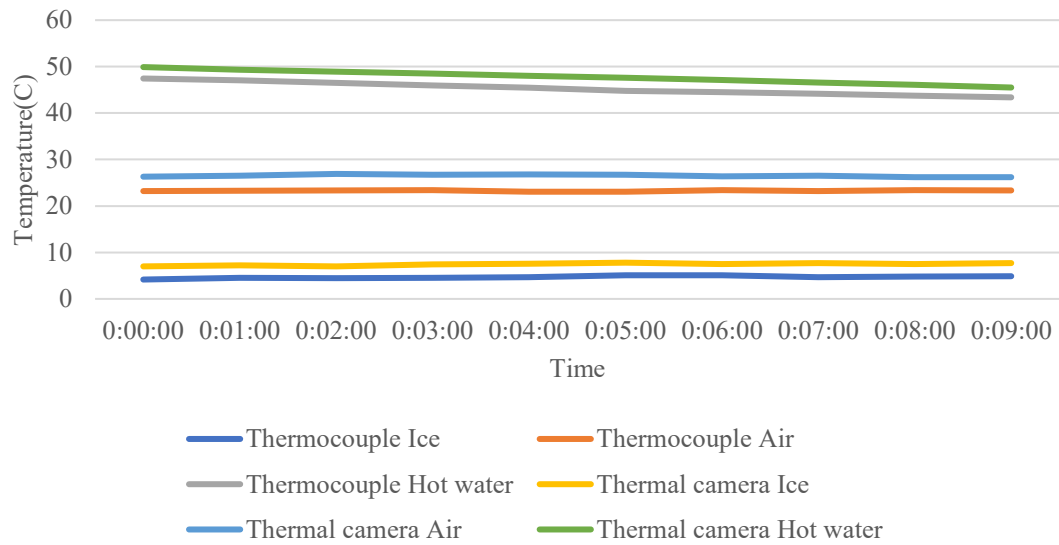


Figure 6-6 Comparing thermal camera and thermocouple measurements over time.

The results of the experiment, illustrated in Figure 6-6, reveal differences between the thermal camera and thermocouple measurements. For hot water, the thermocouple recorded higher temperatures than the thermal camera throughout the test. Although both measurements showed a gradual decline due to natural heat dissipation, the thermal camera readings tended to be lower, likely due to differences in emissivity settings and the way infrared sensors detect radiative heat. The air temperature readings displayed closer agreement between the two devices, with only minor fluctuations observed in the thermal camera data.

For the ice condition, the thermal camera measured lower temperatures compared to the thermocouple. This discrepancy is primarily attributed to the reflective nature of the ice bottle, which affects how infrared radiation is detected. Unlike the thermocouple, which provides a direct-contact measurement, the thermal camera captures infrared radiation that may be influenced by emissivity settings and reflections from nearby surfaces.

Several factors contribute to the observed discrepancies between the two measurement methods. First, emissivity and surface reflectivity play a crucial role in thermal imaging accuracy, particularly for materials with low emissivity, such as water and ice. The

thermal camera relies on detecting emitted infrared radiation, which may vary based on the object's surface properties, leading to potential underestimations. In contrast, the thermocouple provides direct-contact measurements that are not affected by emissivity variations. Additionally, the field of view and measurement approach differ between the two methods. The thermocouple captures temperature from a single precise point, whereas the thermal camera averages temperature values across a broader detection area.

It can be concluded that the FLIR ONE Pro thermal camera does not provide precise absolute temperature values but rather captures temperature differences and heat distribution patterns. From the calibration experiment, slight discrepancies were observed between the thermal camera and the thermocouple. These deviations suggest that thermal imaging is more effective in detecting relative temperature variations rather than providing exact temperature readings.

This limitation is particularly relevant in thermal comfort prediction, where extracting raw temperature values from thermal cameras may introduce inaccuracies due to emissivity variations, sensor calibration, and environmental influences. Instead of relying on absolute temperature readings, this research employs thermal images as direct input for deep learning-based thermal comfort prediction. By using image-based analysis rather than numerical temperature values, the model leverages thermal patterns, ensuring a more adaptable approach.

6.2.4 Deep learning model

The thermal comfort prediction method in this study uses a vision-based deep learning approach, akin to an object detection problem in computer science. In traditional object detection, the objective is to identify and locate objects within an image. Similarly, this method applies to thermal comfort prediction by treating different comfort levels as categories to classify rather than physical objects to locate. This transforms thermal

comfort prediction into a classification task, making it more adaptable to dynamic indoor environments and individual variability.

This study utilises YOLOv8 (ultralytics, 2023), a state-of-the-art single-shot object detection model known for its speed and accuracy in real-time applications. YOLO-based models have been extensively used in tasks such as occupancy detection (Zhang et al., 2024) and object recognition (Bakana et al., 2024). They predict bounding boxes and class probabilities in a single forward pass through the network, offering high efficiency, particularly for real-time scenarios. Unlike traditional two-stage detectors, YOLO-based models avoid the computational overhead of region proposal networks. YOLOv8, the most recent version, features an enhanced architecture, improved feature extraction, and optimised post-processing techniques, delivering superior detection performance compared to its predecessors (Paszke et al., 2019).

As illustrated in Figure 6-7, YOLOv8's architecture comprises four main components: a backbone network for feature extraction, a neck for feature aggregation, a head for final predictions, and a loss function to optimise model performance. The backbone uses convolutional layers to extract hierarchical features from input thermal images, capturing key patterns in thermal distribution across the body. The neck employs a path aggregation network (PAN) to combine low-level and high-level features, enhancing the model's ability to detect subtle variations in thermal data that correspond to different comfort levels. The head generates predictions for each thermal comfort class, enabling precise classification. The loss function minimises the error between predicted and actual thermal comfort levels, ensuring the model's accuracy.

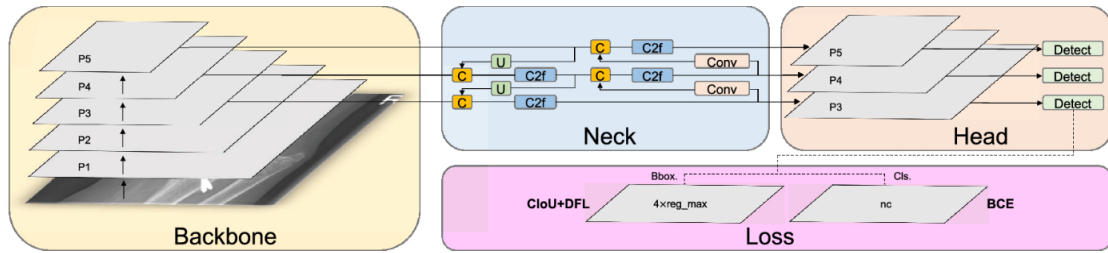


Figure 6-7 The architecture of the YOLOv8 algorithm, which is divided into four parts, including backbone, neck, head, and loss (Ju and Cai, 2023).

In this study, the deep learning model for thermal comfort detection was trained using PyTorch 2.0 (Paszke et al., 2019), a widely recognised open-source deep learning framework. The training process was implemented in Google Colab, leveraging access to an NVIDIA T4 Tensor Core GPU with 2,560 CUDA cores and 16 GB of memory. A workstation running Ubuntu 20.04 with GPU acceleration served as the operating system for the virtual machine, while Python was utilised for coding and implementation.

The dataset for the deep learning model was created from thermal images collected during the experiments. These images were labelled with corresponding TSVs ranging from "-3" (cold) to "+3" (hot). The cooling phase provided the data for training, while video recordings from the heating phase were reserved for validation. Each TSV category represented a distinct class in the deep learning model, resulting in a maximum of seven categories. To ensure sufficient representation, approximately 100 images per category were included in the dataset for each experiment, adhering to YOLO's guidelines (Jocher, 2023).

As the duration of thermal conditions (ranging from cool to hot) varied across subjects, the video recordings were divided into segments of different time intervals. This approach ensured that the number of thermal images within each category remained balanced, despite individual differences in how quickly subjects transitioned through the thermal sensations. By segmenting the videos based on thermal comfort phases, a

uniform dataset was generated. This dataset served as the foundation for training the deep learning model using thermal images.

All images are labelled with bounding boxes by the LabelImg annotation tool (Tzutalin, 2018) and are displayed in a resolution of 1440×1080 pixels. The subject is separated from the background and focuses on the thermal patterns and temperature variations of the subject's body, excluding irrelevant background information. The dataset created contains clean, focused thermal images of the person, which simplifies the input data, removing distractions and allowing the model to concentrate on heat distribution on the subject's body, which is directly related to thermal sensation (Choi and Loftness, 2012). The image annotations generated were saved in a .txt file format for the input of the YOLO algorithm. 70% of the picture datasets are used for training, 20% are used for validation, and the remaining portion is used for testing. All datasets are uploaded and available at Roboflow, a web-based application for objection detection datasets (Roboflow, 2023). Figure 6-8 shows two examples of subject 3 and subject 7 of the images and categories gathered and labelled in the dataset.

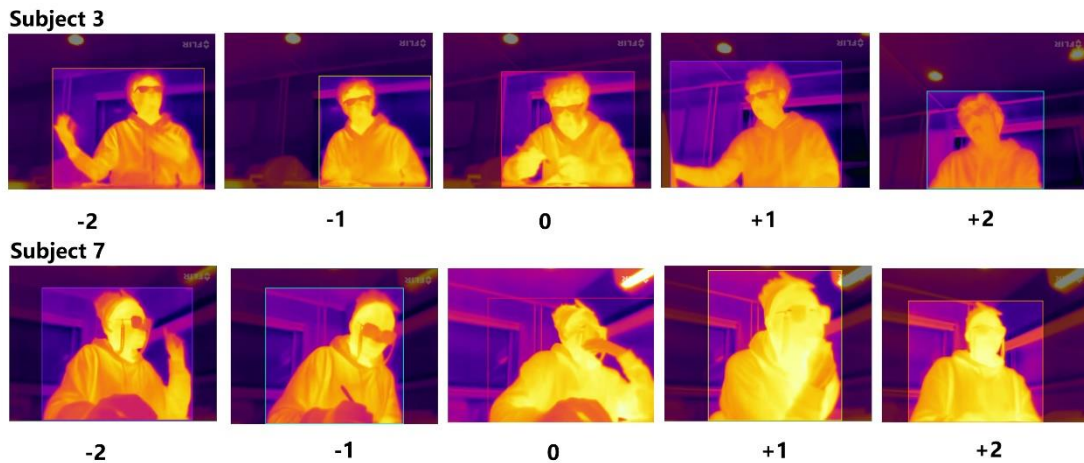


Figure 6-8 Example pictures in the dataset of different classification for subject 3 and subject 7.

Each model for the corresponding subject was trained to establish an individual thermal comfort model using thermal images and associated TSV labels. This approach allowed the model to learn unique thermal patterns and temperature variations specific to each subject, enabling accurate detection of their thermal comfort levels. All models were fully trained, with detailed information on the training datasets provided in Table 6-2. Training continued until no further improvement was observed or when the loss function consistently dropped below a predefined threshold. This process ensured efficient training by avoiding both underfitting (where the model fails to learn enough from the data) and overfitting (where the model learns excessive noise from the training data). To evaluate the performance of the models, the Mean Average Precision (mAP) was measured at an Intersection over Union (IoU) threshold of 0.5 for each experiment, using both the training and validation datasets, as shown in Table 6-2.

Table 6-2 Detailed information about the deep learning model for each subject.

| Subject | Category | Images | Training time | Epochs | mAP ⁵⁰ (%) (Train) | mAP ⁵⁰ (%) (Validate) |
|---------|-----------------------------|--------|---------------|--------|-------------------------------|----------------------------------|
| 1 | "-2", "-1", "0", "+1" | 385 | 1.4h | 259 | 0.94 | 0.94 |
| 2 | "-1", "0", "+1", "+2" | 367 | 1.03h | 199 | 0.98 | 0.81 |
| 3 | "-2", "-1", "0", "+1", "+2" | 415 | 1.45h | 247 | 0.98 | 0.81 |
| 4 | "-1", "0", "+1", "+2" | 472 | 1.47h | 206 | 0.99 | 0.99 |
| 5 | "-1", "0", "+1", "+2" | 360 | 1.07h | 215 | 0.98 | 0.98 |
| 6 | "-1", "0", "+1" | 276 | 0.79h | 214 | 0.97 | 0.97 |
| 7 | "-2", "-1", "0", "+1", "+2" | 608 | 1.15h | 271 | 0.98 | 0.78 |
| 8 | "-2", "-1", "+1", "+2" | 398 | 1.64h | 300 | 0.99 | 0.85 |
| 9 | "-2", "-1", "0", "+1" | 360 | 0.89h | 275 | 0.98 | 0.73 |

| | | | | | | |
|----|--------------------------------|-----|-------|-----|------|-------------|
| 10 | “-2”, “-1”, “+1”, “+2” | 488 | 1.88h | 269 | 0.99 | 0.99 |
| 11 | “-1”, “0”, “+1”, “+2” | 488 | 1.63h | 169 | 0.99 | 0.99 |
| 12 | “-2”, “-1”, “0”, “+2” | 488 | 1.18h | 161 | 0.98 | 0.86 |
| 13 | “-1”, “0”, “+1”, “+2”, “+3” | 608 | 0.97h | 226 | 0.97 | 0.76 |
| 14 | “-2”, “-1”, “0”, “+2” | 488 | 1.15h | 149 | 0.99 | 0.86 |

All models for the corresponding subjects demonstrated good performance, with validation mAP values ranging from 0.73 to 0.99. Subjects 4, 10, and 11 achieved the highest validation mAP of 0.99, highlighting the model’s ability to accurately capture the thermal comfort levels of the corresponding subjects. The performance of these models will be further discussed in subsequent sections.

To evaluate the performance of the deep learning models, Accuracy was used as one of the primary metrics for classification performance (Sokolova et al., 2006). This metric is calculated as the ratio of correctly classified samples to the total number of samples, as defined in the equations below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5 - 5)$$

$$Precision = \frac{TP}{TP + FP} \quad (5 - 6)$$

$$Recall = \frac{TP}{TP + FN} \quad (5 - 7)$$

$$F1\ score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (5 - 8)$$

True Positive (TP) indicates the correct results of thermal level were predicted, False Positive (FP) stands for the incorrect results of thermal level were predicted, False Negative (FN) represents the correct results of thermal level were not predicted, and True Negative (TN) illustrates the correct results of thermal level of the undesired conditions were predicted. A confusion matrix, which compares actual target values with

predicted values, includes counts of true positives (correctly predicted positive cases), true negatives (correctly predicted negative cases), false positives (negative cases incorrectly predicted as positive), and false negatives (positive cases incorrectly predicted as negative).

The normalised confusion matrix transforms raw counts into proportions or percentages, facilitating performance comparison across different classes, where higher values indicate better accuracy. Across all models for the 14 subjects, the classes “-2” and “-1” consistently perform well, often achieving accuracy values approaching or reaching 1.00. This indicates that the model is highly reliable in capturing cold-related thermal comfort states. However, the performance for the warm class shows greater variability, particularly for the neutral class (“0”). The model appears to struggle with maintaining clear distinctions between neutral comfort and warmer states, especially for subjects whose perception of warmth deviates from the average.

The experiments conducted during winter likely influenced the model's performance. In cooler indoor environments, heating systems typically aim to maintain conditions close to neutral or slightly warm, aligning with seasonal expectations. As a result, subjects tend to be more sensitive and accurate in reporting cold discomfort, whereas their perception of warmer conditions may be less pronounced. This seasonal influence has been noted in other research studies, as highlighted in (Qiao and Yan, 2022) and (Fang et al., 2022).

In the intra-subject experiments, some subjects, such as those represented in the confusion matrices for Subjects 1 and 3, demonstrate high accuracy across most classes. However, other subjects exhibit more frequent misclassifications. Appendix.B provides the normalised confusion matrix for 14 models, offering detailed insights into model performance. These results suggest that the model may be overfitting to certain subjects

or that individual differences, such as environmental adaptation, are influencing its accuracy.

For the cross-subject training phase, three experiments were conducted using different subsets of participant data. The results of these experiments are summarised in Table 6-3, which details the training data configuration, the number of images used, the training time, the number of epochs, and the mAP at 50% IoU for both training and validation datasets.

Table 6-3 The training details for the multi-people dataset.

| Experiment | Subject | Images | Training time | Epochs | mAP⁵⁰ (%) (Train) | mAP⁵⁰ (%) (Validate) |
|-------------------|----------------|---------------|----------------------|---------------|---|--|
| 1 | 1-4 and 7-14 | 4977 | 14.69h | 194 | 0.98 | 0.98 |
| 2 | 1-8 and 11-14 | 5027 | 15.13h | 206 | 0.98 | 0.97 |
| 3 | 1-10 and 13-14 | 4667 | 13.45h | 193 | 0.98 | 0.97 |

In the first experiment, the model was trained with 4,977 images over 14.69 hours, distributed across 194 epochs, achieving a mean Average Precision (mAP50) of 98% for both the training and validation sets. This indicates robust model training and excellent generalisation to unseen validation data from Subjects 5 and 6. The second experiment involved a slightly larger dataset of 5,027 images and a marginally longer training duration of 15.13 hours across 206 epochs. While the training accuracy remained high at 98%, there was a slight decrease in validation accuracy to 97%, tested on Subjects 9 and 10. This suggests that while the model effectively learned from a broader range of training data, slight discrepancies in subject-specific thermal responses may affect its universal applicability. In the third experiment, the model was trained using 4,667 images for 13.45 hours over 193 epochs, achieving a training accuracy of 98% and a validation accuracy of 97% when evaluated against Subjects 11 and 12.

In addition to these metrics, a normalised confusion matrix was generated for the training and validation data of the cross-subject model. Appendix.C presents the normalised confusion matrix for all cross-subject models with different subject data, providing a detailed breakdown of the model's performance across different thermal comfort classes. Similar to the intra-subject test, the warm classes exhibit slight confusion compared to the colder classes. This could be attributed to humans generally being more sensitive to cold (Yamazaki et al., 2023). Furthermore, as the experiments were conducted in winter, participants may have expected cooler temperatures and underreported sensations of warmth (Qiao and Yan, 2022).

6.3 Results and discussions

6.3.1 Comparison between PMV and TSV

The most common method for predicting thermal comfort is the PMV, which estimates the thermal comfort level of a group of individuals based on a seven-point thermal sensation scale (Fanger, 1970b). However, actual occupant comfort is subjective, varies among individuals, and is influenced by various factors such as building structure and indoor and outdoor thermal conditions. In this study, three groups of data were collected during the case study to facilitate comparisons: indoor environmental data for PMV calculation (including indoor air temperature, globe temperature, relative humidity, and airspeed); thermal video to test the proposed deep learning method; and TSVs obtained from the occupants. To ensure consistency in clothing insulation and metabolic rate, all subjects were instructed to wear a single layer of clothing throughout the experiment.

The PMV method is widely adopted by researchers and practitioners globally and is included in several national building codes and international standards, such as ASHRAE 55–2023 (Standard, 1992), EN 16798–1:2022 (CEN, 2019), and ISO 7730:2005 (AC08024865, 2005). According to Fanger's thermal comfort equation [63],

human thermal comfort can be determined using four environmental factors—air temperature (T_a), relative humidity, mean radiant temperature (T_{mr}), and air velocity (V)—along with two personal factors: clothing insulation (I_{cl}) and metabolic rate (M). Equations (1), (2), (3), and (4) detail the calculation process.

$$PMV = [0.303 \exp(-0.036M) + 0.0275] \{ (M - W) - 3.96 \cdot 10^{-8} \cdot f_{cl} [(T_{cl} + 273)^4 - (T_{mr} + 273)^4] - f_{cl} \cdot h_c (T_{cl} - T_a) - C_1 - C_2 \} (5 - 1)$$

Where

$$C_1 = 3.05 \cdot 10^{-3} [5733 - 6.99(M - W) - P_a]$$

$$C_2 = 0.42[(M - W) - 58.15] - 1.7 \cdot 10^{-5} M (5867 - P_a) - 0.0014 \cdot M (34 - T_a)$$

Where

W represents heat generated by external work, W/m^2 ;

P_a is vapour pressure in ambient air, kPa;

f_{cl} and h_c are clothing factor and convective heat transfer coefficient with units of $W/(m^2 \cdot K)$;

T_{cl} is the temperature of the clothing surface, °C.

And,

$$T_{cl} = 35.7 - 0.0275 \cdot (M - W) - 0.155 \cdot I_{cl} \cdot [(M - W)] - C_1 - C_2 \quad (5 - 2)$$

And

$$h_c = \max \left[\frac{[2.38(T_{cl} - T_a)]^{0.25}}{12.1 \cdot \sqrt{V}} \right]$$

$$f_{cl} = \begin{cases} 1.0 + 0.2 \cdot I_{cl} & \text{if } I_{cl} < 0.5clo \\ 1.05 + 0.1 \cdot I_{cl} & \text{if } I_{cl} > 0.5clo \end{cases} \quad (5 - 3)$$

The mean radiant temperature was determined with the measured data at the

site.

$$T_{mr} = \left[(T_{gt} + 273.15)^4 + \frac{1.1 \cdot 10^8 \cdot V^{0.6}}{\varepsilon \cdot D^{0.4}} (T_{gt} - T_a) \right]^{\frac{1}{4}} - 273.15 \quad (5 - 4)$$

where

D, the diameter of the globe was 0.05 mm;

ε , the emissivity of the surface was 0.9;

T_{gt} is the globe temperature, °C.

The PMV was obtained using the CBE thermal comfort tool (Tartarini et al., 2020), a tool for thermal comfort indices calculation and visualisation for the standard ASHRAE 55–2023, EN 16798–1:2022 and ISO 7730:2005 with the equations above. Air temperature (T_a), relative humidity (RH), mean radiant temperature (T_{mr}) and air velocity (V) were collected from the case study experiment and used for PMV calculation. The clothing insulation I_{cl} and metabolic rate M were consistent since the subjects were asked to wear one layer of cloth in all tests.

Figure 6-9 illustrates the changes in PMV values over time for 14 experiments. At the start of the experiments, most subjects reported experiencing cold discomfort due to the AC being off—an expected condition during winter. Once the AC was set to heating mode (temperature setpoint: 30°C), PMV values generally trended toward neutrality or slightly positive comfort levels. However, the time taken for each subject to achieve this state varied. For instance, subjects like 1 and 5 transitioned to more comfortable PMV levels relatively quickly, while others, such as 5 and 6, required significantly longer to reach neutrality. This variation in response time can be attributed to factors such as the effectiveness of the heating system and the individual thermal sensations of each participant. Participants who achieved comfort more quickly may exhibit a higher

sensitivity to incremental increases in temperature, while others may have different physiological or psychological thresholds for perceiving thermal comfort.

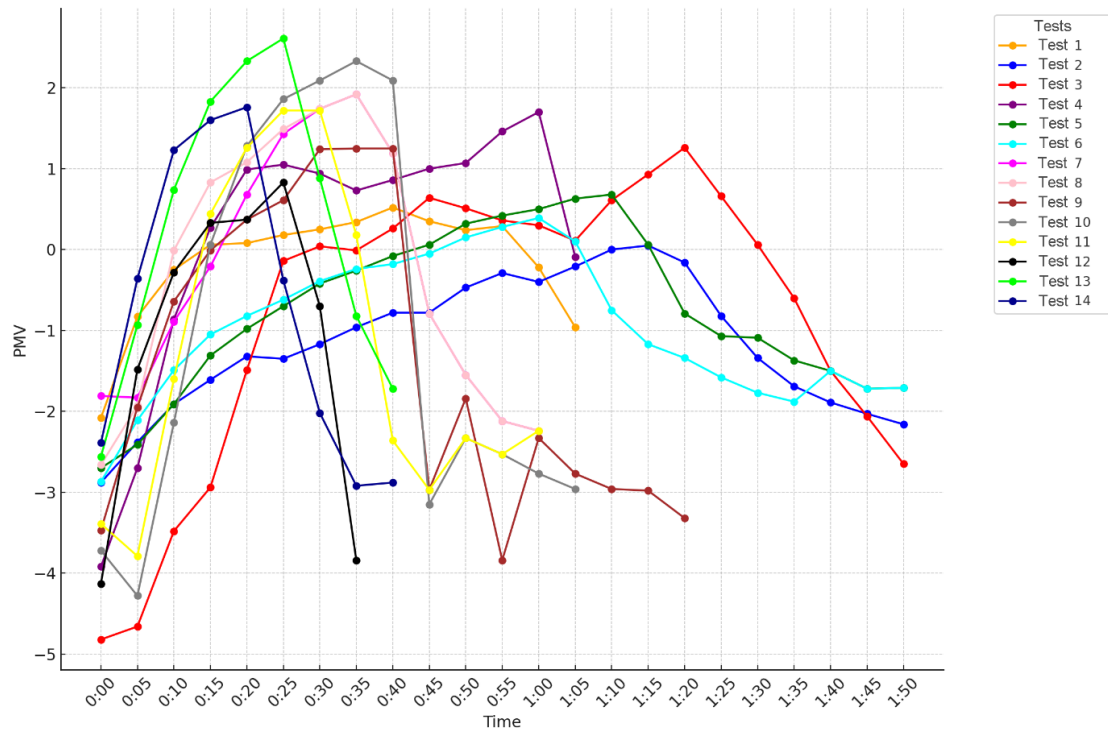


Figure 6-9 The PMV values over time in 14 tests.

The PMV method is designed to describe the overall thermal conditions of space, whereas TSV is targeted at capturing individual perceptions of thermal comfort. In this study, TSVs were recorded every five minutes through questionnaires, while PMVs were calculated using environmental data collected at the same intervals. This dual approach provides an objective prediction of thermal comfort based on physical parameters while also capturing subjective individual responses.

Figure 6-10 illustrates the relationship between PMV and TSV across 14 experiments, revealing a general correlation between the two measures. However, notable deviations were observed, with some subjects reporting stronger sensations of cold or warmth (TSV) than those suggested by the PMV values. This disparity indicates that while PMV

is a valuable tool for predicting thermal comfort, it does not always fully reflect individual perceptions—a limitation noted in previous studies (Laouadi, 2022).

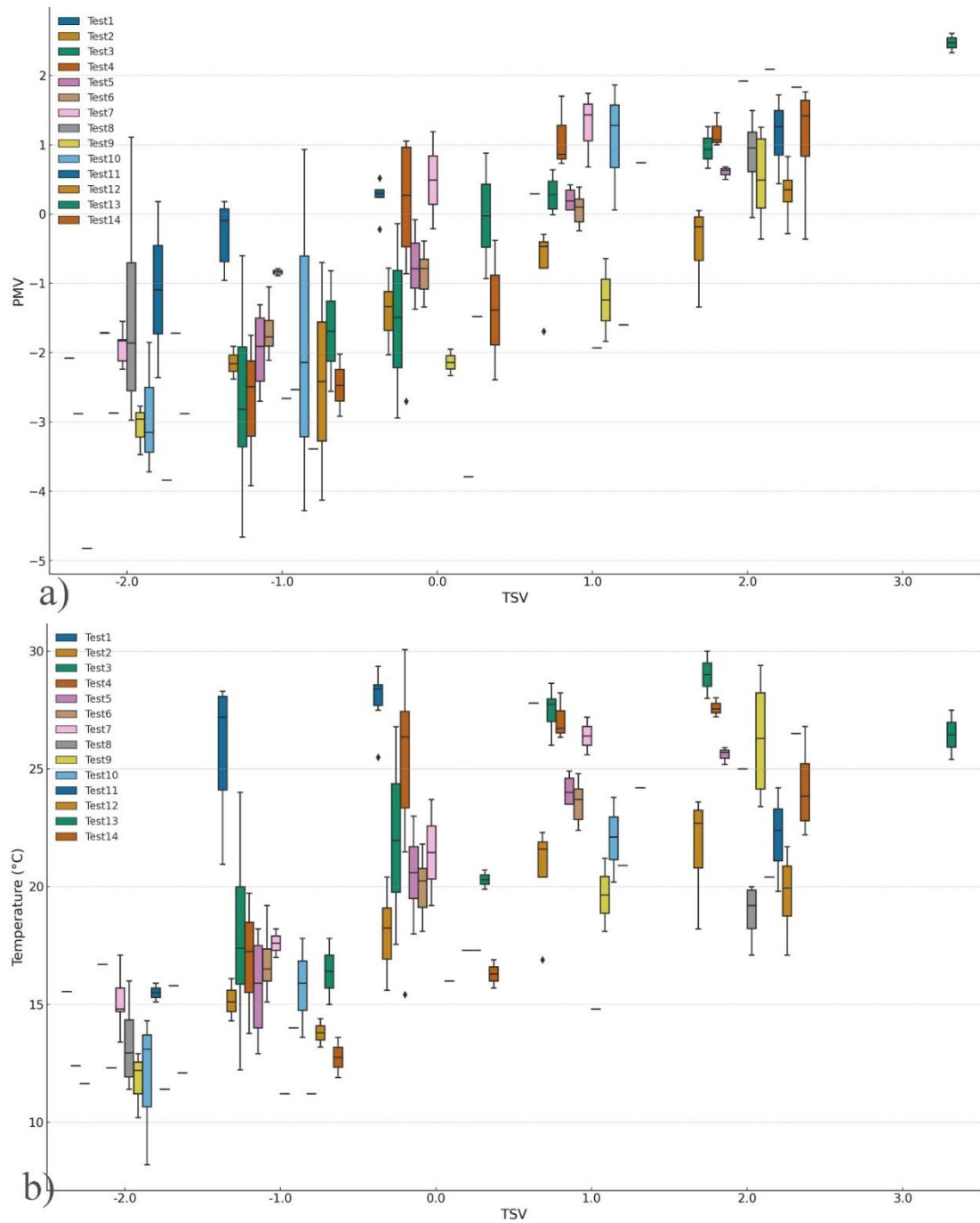


Figure 6-10 Comparison of a) TSV against temperature and b) TSV against PMV in 14 experiments.

Despite the overall alignment, there are clear discrepancies between PMV and TSV across subjects. It is worth noting that PMV is not designed for individual-level predictions but is being used here for comparison purposes. For example, Subjects 1 and 5 align fairly with PMV and TSV as temperatures increase. However, Subject 1's TSV is slightly more negative than the corresponding PMV values at certain points while the TSV for Subject 5 indicates slight warmth than the PMV. Subject 3 shows a more pronounced divergence between PMV and TSV, especially at lower temperatures, this person might have a slower physiological adaptation to changes in environmental conditions, or possibly a preference for warmer environments, which the PMV model doesn't fully capture. Subjects 9 and 10's TSV remains lower when PMV predicts neutral or slightly warm conditions, indicating that they perceive more discomfort than expected.

The comparison between PMV and TSV highlights the limitations of the PMV model in accurately predicting individual thermal comfort, especially in cases where there are significant variations between subjective sensations and objective predictions. While PMV can serve as a general predictor of comfort, it often fails to account for the complexities of individual thermal perceptions, particularly at the extremes of the temperature scale (Yau and Chew, 2012).

6.3.2 Deep learning prediction of thermal comfort based on the intra-subject dataset

To evaluate the effectiveness of the deep learning model in detecting indoor thermal comfort from thermal images, two tests were conducted: one using intra-subject datasets for individuals and another using cross-subject datasets for multiple people.

The intra-subject tests aim to assess the accuracy of the model in predicting individuals' unique thermal comfort responses, with both the training dataset and validation video

coming from the same subject. In these experiments, deep learning detection results were collected every 1 minute, and their 5-minute averages were compiled to provide a clearer picture of the model's performance over time. The differences between PMV, TSV, and deep learning detection results over the validation video for the 14 subjects are visualised in Figure 6-11. According to ASHRAE 55–2023 (Standard, 1992), thermal neutrality is defined as $-0.5 < \text{PMV} < +0.5$. In this study, the range $-1 < \text{PMV} < +1$ was classified as Neutral, while -2 was defined as Cold and $+2$ as Hot, as the extreme values of -3 and $+3$ were rarely observed during the experiment. Figure 6-11 illustrates the detailed results of the personal vision-based deep learning models, along with the corresponding TSV values from the survey and PMV values calculated from environmental data across 14 tests.

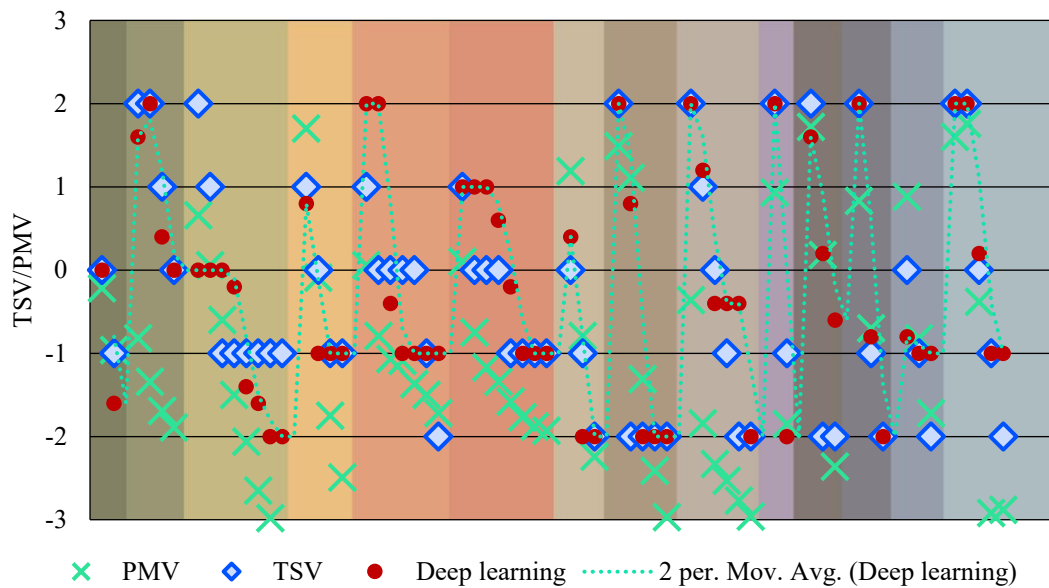


Figure 6-11 Comparison between TSV, PMV and deep-learning model results for 14 subjects.

The duration of the validation video varies between 10 and 40 minutes due to differences in how quickly the room's heating or cooling system reaches stable conditions and individual differences in how subjects adapt to temperature changes, with some

requiring more time to feel comfortable or uncomfortable. For most subjects, PMV aligns well with TSV when temperatures are close to neutral conditions (typically between 20–26°C). For example, in Subject 1 at 16:30, the PMV is -0.22, which corresponds closely with a TSV of 0, indicating that the PMV model accurately reflects the subject's neutral thermal sensation. Similarly, for Subject 9 at 14:05, the PMV of -0.36 aligns well with a TSV of 2, suggesting that the model captures the subject's perception of slight warmth.

However, TSV values often diverge from PMV under extreme conditions when PMV exceeds 2 or falls below -2, or during sudden temperature changes. For instance, in Subject 2 at 15:40, the PMV of -1.89 corresponds to a TSV of 0, suggesting that PMV underestimates the subject's thermal tolerance. Similarly, Subject 3 at 17:11 has a PMV of -3.32 but a TSV of -1, indicating that the subject may not experience the extreme discomfort predicted by PMV. The deep learning method provides additional insights, with some instances where its predictions closely match TSV, even when PMV does not. For example, in Subject 7 during the first 5 minutes, the deep learning model predicts a value of 0.4, which is closer to the TSV of 0, whereas PMV predicts 1.19. However, in other cases, such as Subject 5 at 15:50 (PMV of -0.79, TSV of 0, deep learning prediction of 2), the deep learning model diverges significantly, potentially indicating overfitting or sensitivity to noise in the data.

As an example, Figure 6-12 and

Video 6-1 present the deep learning model validation result for subject 2 from 15:30 to 15:35. The TSV from the survey was 2 while PMV was -1.34 from 15:30 to 15:35 and the deep learning model correctly predicted a TSV of 2 for most of the 5 mins. The video demonstrates the model's ability to provide real-time thermal comfort detection. But during the interval between 15:31 and 15:32 the model's results briefly fluctuated, predicting 0 and 1 instead of the actual TSV of 2.

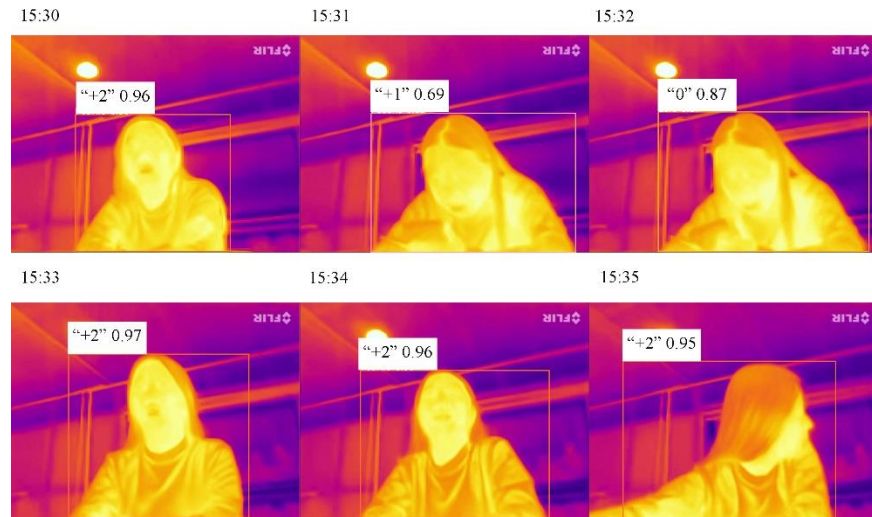


Figure 6-12 The screenshots of the validation video from 15:30-15:35 for every 1 minute in the experiment of Subject 2. The first number indicates the TSV, and the second number is the PMV.



Video 6-1 The video from 15:30-15:35 (accelerated) in the experiment for subject 2 in which the TSV was 2, PMV was -1.34 and deep learning result varied from 0 to 2. The playable video is available at <https://youtube.com/shorts/4sNOew4Dqo0?feature=share>

In this study, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) (Chai and Draxler, 2014) were used to evaluate the accuracy of the thermal comfort prediction models, including both the PMV model and the deep learning model, in comparison to the actual TSV from the subjects. These metrics were chosen due to their complementary abilities to assess model performance in a way that is directly applicable to real-time thermal comfort prediction.

Mean Absolute Error (MAE) captures the average magnitude of errors between model predictions and actual TSV values, representing how far the model's predictions are, on average, from the true thermal comfort values perceived by the occupants. The formula for MAE is:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Predicted_i - Actual_i| \quad (5 - 5)$$

Where n is the total number of observations, $Predicted_i$ is the model's prediction for the i -th observation, and $Actual_i$ represents the actual TSV. In this study, a lower MAE indicates that the model's predictions align more closely, on average, with the occupants' reported comfort levels, suggesting reliable predictive accuracy in real-time applications.

Root Mean Square Error (RMSE) places more emphasis on larger errors by squaring each error before averaging, which makes it sensitive to significant deviations. In the thermal comfort detection model, where large errors can lead to uncomfortable conditions and inefficient HVAC performance, RMSE helps to highlight any tendency of the model to produce substantial outliers. By penalising larger errors more heavily, RMSE indicates whether the model occasionally produces significant deviations from actual TSVs, which could negatively impact system stability and climate control reliability. The formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Predicted_i - Actual_i)^2} \quad (5 - 6)$$

Table 6-4 The MAE and RMSE of the PMV and deep learning method results for 14 subjects.

| Subject | PMV | | Deep Learning | |
|---------|------|------|---------------|------|
| | MAE | RMSE | MAE | RMSE |
| 1 | 0.13 | 0.16 | 0.3 | 0.42 |
| 2 | 2.68 | 2.73 | 0.25 | 0.36 |
| 3 | 0.8 | 0.88 | 1.2 | 1.29 |
| 4 | 0.76 | 0.91 | 0.3 | 0.51 |
| 5 | 0.97 | 0.98 | 1.1 | 1.24 |
| 6 | 1.04 | 1.07 | 0.65 | 0.77 |
| 7 | 0.55 | 0.71 | 0.47 | 0.62 |
| 8 | 1.44 | 1.86 | 0.93 | 1.62 |
| 9 | 2.51 | 2.52 | 0.2 | 0.26 |
| 10 | 0.96 | 0.97 | 0.5 | 0.71 |
| 11 | 0.94 | 1.29 | 1.33 | 1.52 |
| 12 | 1.1 | 1.27 | 0.07 | 0.12 |
| 13 | 0.53 | 0.64 | 0.4 | 0.57 |
| 14 | 0.32 | 0.33 | 0 | 0 |

Table 6-4 summarises the MAE and RMSE values for both the PMV and deep learning methods across 14 subjects. Overall, the deep learning model demonstrates superior performance on average compared to the PMV model. It aligns more closely with actual TSV values, showcasing its adaptability to individual thermal responses. For example, with Subjects 2, 4, 6, 7, 9, 10, 12, and 14, the deep learning model exhibited both lower MAE and RMSE compared to the PMV model. This reflects not only the model's accuracy in aligning with the actual TSV but also its reduced susceptibility to large deviations, as indicated by the RMSE. The relatively small RMSE values across most subjects suggest the deep learning model's ability to maintain consistent predictions, a critical requirement for real-time HVAC adjustments, where large prediction errors

could result in discomfort or inefficiency. Conversely, the PMV model, as evidenced in Subjects 7, 8, and 11, struggled with occasional large deviations from the TSV, reflected in higher RMSE values. This highlights the static nature of the PMV method, which is less effective at capturing individual variability. In contrast, the deep learning model provides a more dynamic and accurate alternative, particularly in scenarios where individual thermal responses vary significantly.

This study highlights the promising potential of the deep learning model for thermal comfort detection, particularly in real-time, personalised, and adaptable applications. The model effectively captures specific temperature distributions and body heat patterns without relying on individual environmental sensors or extensive manual input. Furthermore, the deep learning approach is less intrusive and respects privacy by using low-resolution thermal images that reveal only temperature patterns, without identifiable details of occupants. This feature makes it especially advantageous for workplaces or public buildings, where real-time comfort monitoring is needed without compromising privacy.

As an initial study, the thermal images were collected from a single occupant under controlled conditions. However, real-world indoor environments are dynamic, often involving multiple occupants with varying thermal responses, which could affect the model's accuracy. Additionally, the dataset used in this study is relatively small, comprising only 14 subjects, which limits the model's ability to accommodate diverse comfort profiles and environmental conditions. Expanding the dataset to include a broader range of subjects and environmental conditions would enhance the model's training, improving its accuracy and adaptability for more generalised applications.

6.3.3 Deep learning-based detection of thermal comfort based on cross-subject dataset

In the cross-subject phase, datasets from multiple subjects initially used in the intra-subject models were combined to create a comprehensive training set. This approach aimed to develop a more generalised thermal comfort model. Three different training configurations were employed to capture a wide range of thermal responses. In the first configuration, data from subjects 1–4 and 7–14 were used for training, with subjects 5 and 6 reserved for testing. The second configuration combined data from subjects 1–8 and 11–14 for training, leaving subjects 9 and 10 for testing. The third configuration trained the model on data from subjects 1–10 and 13–14, using subjects 11 and 12 for validation. Cross-subject testing on unseen subjects offers insights into the model's potential for real-world thermal comfort prediction, assessing its adaptability and effectiveness across broader applications. The detailed results of these tests are presented in Figure 6-13.

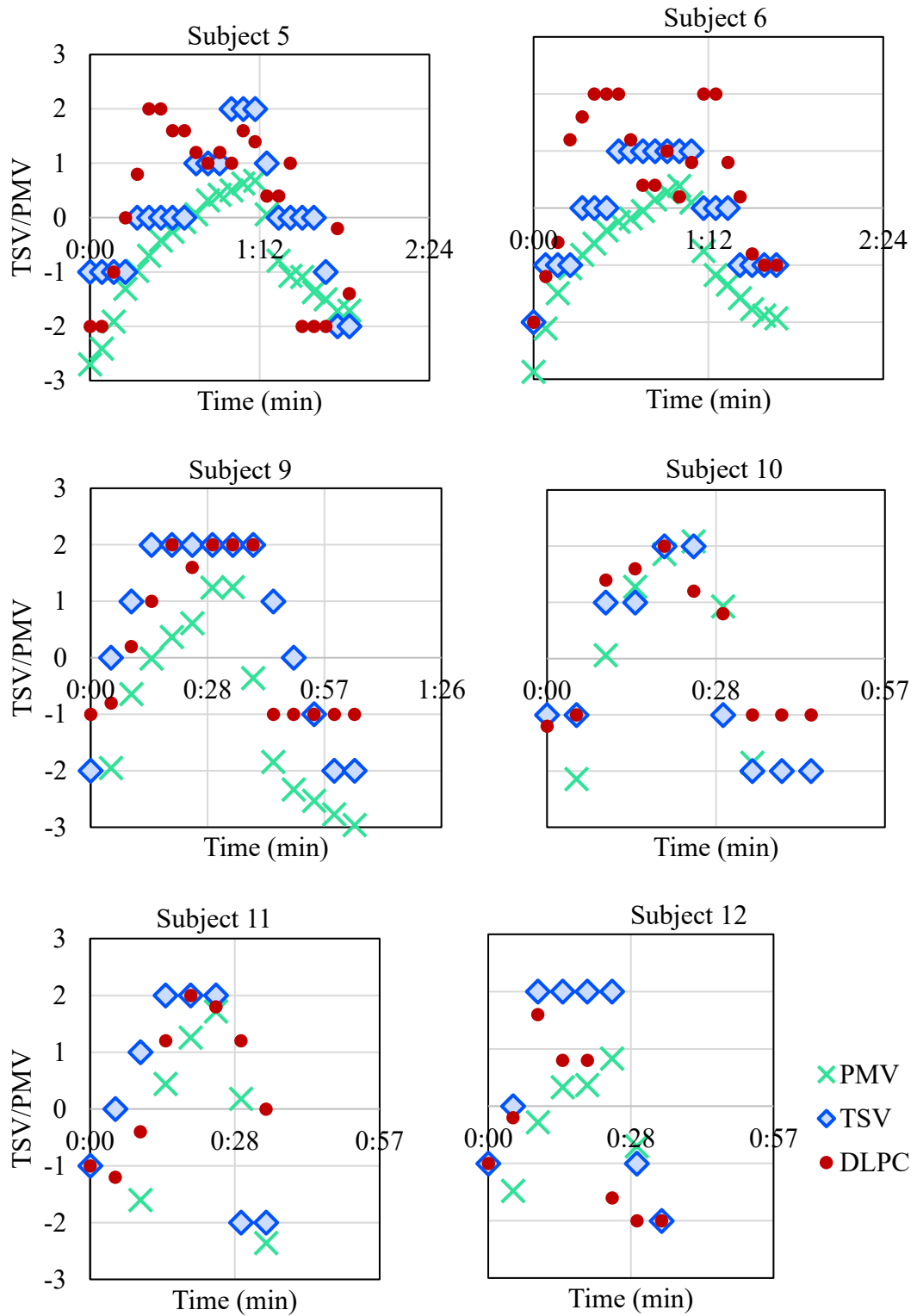


Figure 6-13 The result comparison of the cross-subject models with corresponding TSV and PMV.

The cross-subject model's performance exhibits variability across different subjects, reflecting both its potential and its limitations. For Subject 9, the deep learning model

achieved an accuracy of 84.62%, while the PMV model only had 15.38% accuracy, demonstrating that the deep learning model can effectively capture complex, individualised thermal comfort responses under certain conditions. In contrast, for Subject 10, the deep learning model's accuracy dropped to 50%, while the PMV model outperformed it with 80% accuracy. The deep learning model relies on patterns from its training data and can struggle when faced with unseen individuals whose thermal comfort responses diverge from those patterns (Zhang et al., 2019). In such cases, the PMV model can perform better due to its nature of generalised assumptions, providing stable but less personalised predictions.

In addition, Subject 5's accuracy was 43.48% for the deep learning model, less than the PMV model's 69.56% accuracy. Subject 6 showed a higher accuracy (76.2%) for the deep learning model compared to the PMV model, which had the same performance (76.2%). For both Subjects 11 and 12, the deep learning model performed better than the PMV model, the deep learning accuracy reached 62.5% for Subject 11 and 70% for 12 while PMV only had 12.5% and 20%. The deep learning model likely captured individual-specific thermal response patterns and unique environmental interactions, enabling higher accuracy. In contrast, the PMV model's generalised assumptions failed to account for these subjects' distinct variations, leading to poor performance.

The results suggest that the limited diversity of the training dataset may affect the cross-subject model's ability. Subjects may differ in metabolic rates, clothing insulation, or even thermal sensitivity, which impact how they experience thermal comfort. The deep learning model can identify complex, subject-specific features. Still, this strength can lead to overfitting, where it performs well on familiar data but struggles with new data from different individuals. This explains why the model performed well for some intra-subject models but failed to generalise effectively across others.

Moreover, these results reflect the trade-off between personalised and generalised models. The PMV model, while static and less adaptive, sometimes achieves higher accuracy, as seen with Subject 10, due to its broad applicability based on standardised equations. However, the deep learning model demonstrates better adaptability when trained on diverse data, as with Subject 9. This variability indicates that the model's effectiveness is tied to the characteristics of the individuals included in the training dataset.

As an example of the detail performance of the deep learning model, Figure 6-14 illustrates the progression of the experiment of Subject 5 from cold to hot and then back to cold, showcasing the changes throughout the test as an example. While these images visually appear quite different between the cold and hot phases, the deep learning model effectively captures the comfort level by extracting detailed information from the thermal images that may not be immediately apparent to the human eye. It allows the deep learning model to capture subtle changes in temperature patterns and body heat distribution critical to assessing thermal comfort but may be missed through traditional visual analysis or simpler models.

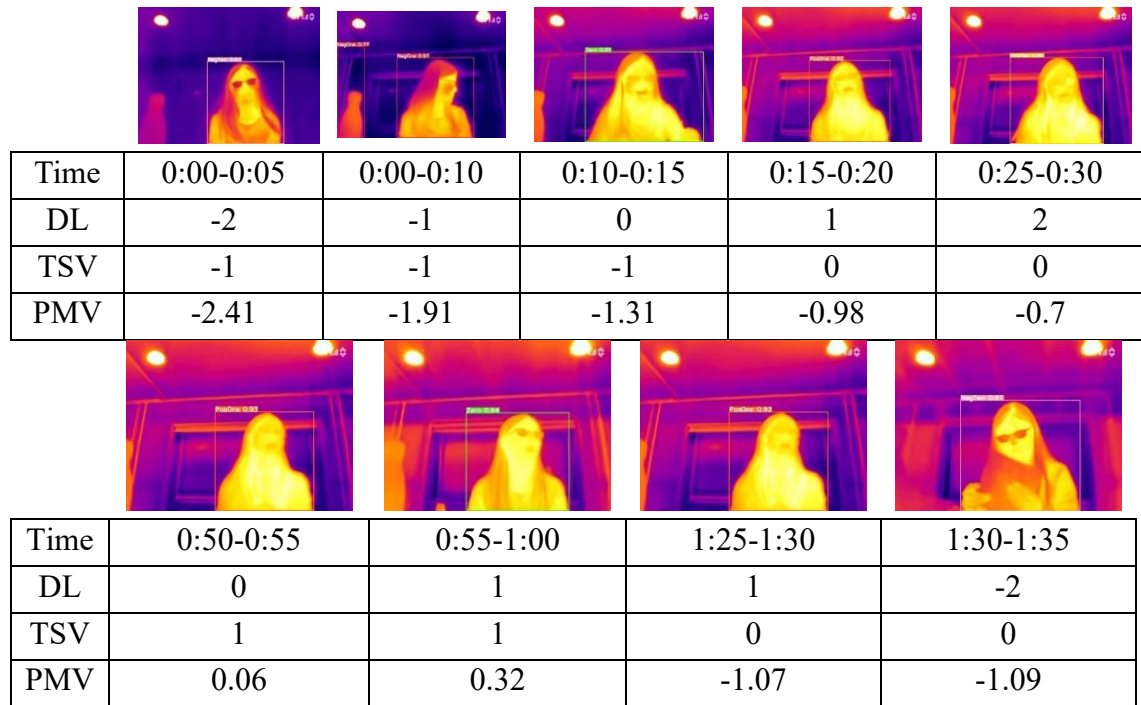


Figure 6-14 The screenshots of results in a cross-subject test for Subject 5 with the dataset from Subjects 1 to 4 and 7 to 14.

Over the past few years, some studies that perform well have been employed to forecast the thermal comfort of occupation (Gao et al., 2020b). The majority rely on complex sensing platforms that use many devices and advanced technical abilities for feature selection and data processing (Aryal and Becerik-Gerber, 2019). The quality of the features that are obtained to train the model typically affects accuracy, requiring extensive data engineering skills (Aryal and Becerik-Gerber, 2020). This study, while initial, demonstrates potential by employing a more straightforward and cost-effective approach using a single thermal camera and a deep learning model. Further development should focus on enhancing the deep learning model by expanding the training dataset to include a broader range of subjects and environmental conditions, enabling the model to generalise more effectively across diverse occupants.

This chapter demonstrates the potential of a vision-based deep learning model for thermal comfort prediction but also uncovers several limitations. In intra-subject tests, while the deep learning model consistently outperformed the PMV model, its accuracy

varied across individual subjects, suggesting a degree of overfitting due to the limited dataset. This variability highlights the need for more personalized modelling approaches and a broader representation of thermal comfort patterns in the training data. In cross-subject tests, the model's performance was inconsistent, with better performance for subjects like 11 and 12. Although the deep learning model performed better than the PMV model for these subjects, its inability to generalise effectively across unseen individuals indicates challenges in capturing inter-subject differences, likely due to the narrow range of physiological, behavioural, and environmental conditions present in the dataset. Moreover, the controlled indoor environment during winter further constrained the diversity of conditions encountered, limiting the model's applicability to real-world scenarios that include extreme temperatures, varying humidity, and dynamic outdoor environments. Additionally, the study focused exclusively on single-occupant scenarios, leaving the model's performance in multi-occupant settings unexplored, where complex interactions between occupants and the environment significantly influence thermal comfort. To address these limitations, future research should incorporate larger, more diverse datasets, expand the range of environmental conditions tested, and evaluate multi-occupant dynamics. Furthermore, strategies such as transfer learning, model personalization, or incorporating adaptive mechanisms should be explored to enhance the model's robustness, adaptability, and scalability for broader, real-world applications.

6.4 Summary

This chapter explored the deep learning model for real-time thermal comfort prediction based on thermal images with a single-shot detection algorithm, YOLOv8, demonstrating its potential as a promising alternative to traditional models like the PMV. The results showed that while the intra-subject deep learning model generally achieved better average performance in terms of predictive accuracy and adaptability, its cross-subject performance highlighted challenges related to generalisation.

Experiments were conducted with 14 individual subjects in a controlled indoor environment, resulting in a dataset comprising 4,977 thermal images and corresponding environmental data points. These were collected to evaluate the model's performance in predicting thermal comfort levels. The result indicated that the deep learning model outperformed the PMV model in several cases, particularly during intra-subject analysis, where it achieved an overall accuracy of 68.49%. Specific subjects, such as Subjects 2, 4, and 6, reached near-perfect accuracy, highlighting the model's capability to capture individual thermal responses effectively. However, for the cross-subject model, the deep learning model's accuracy varied, from the best result of 84.6% for Subject 9 to 43.5% for Subject 5 which demonstrated the challenge of generalising to unseen individuals. Additionally, subjects like Subject 10 (50% accuracy for deep learning compared to 80% for PMV) showed that while the deep learning model has the potential to identify complex thermal comfort patterns, its performance can be compromised when applied to new individuals whose responses differ from the training data.

Despite these challenges, the initial method showed advantages, including its ability to directly process thermal images and extract complex information without the need for extensive and intrusive sensor setups. This capability simplifies the data collection process and offers an adaptable solution for real-time HVAC control, enhancing occupant comfort and energy efficiency.

For future work, one promising direction is the collection of a more extensive and diverse dataset that incorporates various demographics and climate zones. By including data from individuals of different ages, genders, occupations, and geographical locations, the model can capture a broader thermal comfort response. Individuals living in different climate zones may exhibit distinct thermal comfort preferences due to varying weather patterns and seasonal changes. By incorporating climate-specific parameters into the model, it can more accurately predict and adjust indoor thermal conditions based on regional climate characteristics. This adaptation would improve the model's

effectiveness across different geographical areas and seasonal variations, ensuring optimal comfort levels year-round.

Importantly, this technology has significant potential to benefit individuals who may not be able to communicate their thermal comfort needs or adjust their environment, such as those with disabilities, the elderly, young children, or people who are sleeping. The ability to automatically adjust the indoor climate to maintain comfort without requiring user input can improve the quality of life for these individuals, ensuring they remain in a comfortable environment. This aspect highlights the model's potential for inclusive design, making it an invaluable tool in creating more accessible and supportive living and working spaces.

7. CONCLUSION AND FUTURE WORK

7.1 Conclusions

This thesis investigated key challenges and opportunities in vision-based occupancy detection and thermal comfort prediction using deep learning methods, with a focus on advancing smart building technologies. Through a series of experiments and analyses, the research explored multiple dimensions of vision-based detection, including algorithm performance, camera modality comparisons, and thermal imaging for thermal comfort prediction.

The evaluation of eight deep learning models, including YOLO variants, SSD, and Faster R-CNN, for real-time occupancy detection revealed differences in their performance based on accuracy, speed, and computational efficiency. YOLOv8x demonstrated the highest accuracy (77%) but with increased inference time, while YOLOv8n achieved a balance between speed and accuracy, making it suitable for dynamic and crowded scenarios. In contrast, models like SSD showed limited performance, struggling with occupant detection in complex or crowded settings. It also highlighted the impact of camera placement and multi-camera setups in addressing occlusion issues, though these approaches increase system complexity and costs. The ability of deep learning models to reduce the gap between predicted and actual energy consumption, with a maximum error reduction of 6.72% compared to conventional methods, underscored their potential for improving demand-driven building energy management. Furthermore, the deep learning models demonstrated the ability to align CO₂ concentration trends more accurately than traditional approaches, highlighting their applicability in real-time energy performance prediction.

The comparison between standard and thermal cameras revealed that both camera types could achieve occupancy detection accuracies of approximately 70% with YOLOv8 and 80% with YOLOv10, given sufficient dataset preparation and diverse training images. In the Same-Video experiment, both cameras achieved approximately 94% accuracy and a mean average precision (mAP) of 0.8 in controlled settings, which demonstrated their maximum potential. Standard cameras performed better in controlled settings with high resolution but faced challenges in privacy-sensitive applications and scenarios involving distractions like portrait images. Thermal cameras, while initially less precise due to lower resolution, offered advantages in privacy protection and low-light conditions. The highest accuracy observed was 88% for normal cameras and 83% for thermal cameras in Cross-Video Experiment 3, indicating that dataset complexity influenced model generalization. The results demonstrated that thermal cameras could effectively decrease issues like visual distractions and residual heat signatures with diverse and targeted training datasets. For instance, overlapping occupants, a major challenge in earlier experiments, were detected in later experiments as dataset diversity improved.

Given the strengths and limitations of both camera types, this research extended the vision-based approach to thermal comfort prediction as an alternative to the Predicted Mean Vote (PMV) model. It demonstrated the potential of thermal cameras to directly process thermal images and predict individual comfort levels in real time. Experiments were conducted with 14 individual subjects in a controlled indoor environment, resulting in a dataset comprising 4,977 thermal images and corresponding environmental data points. The deep learning model achieved an overall accuracy of 68.49% in intra-subject analysis, surpassing the traditional PMV model in several cases. However, when applied across multiple subjects, model accuracy varied from 84.6% for Subject 9 to 43.5% for Subject 5, underscoring the high variability in individual thermal responses and the difficulty in developing a generalised thermal comfort model. These

findings emphasize the challenge of generalization when applying models to unseen subjects. Despite this limitation, the ability to process thermal images directly simplifies data collection and offers an adaptable approach for real-time HVAC control, improving both occupant comfort and energy efficiency.

This thesis has addressed its overarching aim: to explore the application of vision-based deep learning frameworks for improving occupancy prediction and thermal comfort modelling in building environments. Each of the stated objectives has been systematically met. First, a comprehensive review of existing literature on machine learning applications in building systems was conducted, with a particular focus on occupancy detection, indoor air quality, thermal comfort, and energy consumption optimisation. Second, diverse datasets were collected, annotated, and tested—featuring both thermal and RGB imagery in realistic indoor scenarios—to support model development. Third, multiple deep learning algorithms, including SSD, Faster R-CNN, and YOLO variants, were evaluated for their performance in real-time occupancy prediction. Fourth, the study conducted a comparative analysis of standard and thermal cameras, highlighting key trade-offs in accuracy, privacy, and application context. Fifth, the integration of occupancy prediction into energy simulation allowed for the quantification of its impact on heating energy demand and CO₂ levels. Finally, a novel, non-intrusive thermal comfort prediction model was developed using thermal imaging and deep learning techniques, demonstrating the feasibility of personalised comfort modelling. Collectively, these achievements contribute to the development of adaptive, occupant-aware building systems that align with the thesis aim and address existing gaps in the literature.

Here's the key findings of this thesis:

- YOLOv8n provided the best balance between speed and accuracy for real-time occupancy detection in building environments. YOLOv8x achieved the highest

detection accuracy (77%) but had longer inference times, making it less practical for real-time use.

- SSD and Faster R-CNN performed less effectively in complex and crowded indoor settings.
- Diverse datasets significantly improved model generalisation, particularly for overlapping occupants and varied lighting conditions.
- Thermal and RGB cameras both achieved around 94% detection accuracy in controlled environments with sufficient training data.
- Thermal cameras outperformed RGB cameras in privacy-sensitive and low-light conditions, while RGB cameras offered better resolution and detail recognition.
- Deep learning-derived occupancy profiles reduced energy simulation errors by up to 6.72% compared to fixed occupancy schedules.
- CO₂ concentration trends predicted using occupancy-informed profiles were more accurate than traditional static assumptions.
- A thermal comfort model using deep learning and thermal images achieved 68.49% accuracy in intra-subject tests, often outperforming the PMV model.
- Cross-subject thermal comfort prediction varied widely (43.5%–84.6%), confirming strong individual differences in thermal sensation.
- The deep learning model enabled non-intrusive, real-time thermal comfort assessment without the need for wearable sensors.

- Several new datasets were created, including RGB and thermal image sets annotated for occupancy and comfort under varying environmental conditions.
- The research introduced the concept of dynamic thermal comfort using image-based modelling, advancing personalised HVAC control.
- A novel vision-based framework was proposed, integrating detection, modelling, and energy simulation, offering a pathway to intelligent building management systems.

7.2 Contribution to Knowledge

This thesis makes several contributions to the field of vision-based occupancy prediction and thermal comfort modelling, addressing critical gaps identified in the literature and aligning with the study's objectives.

Several new datasets were developed including occupancy prediction in indoor environments, corresponding normal and thermal camera images captured under diverse environmental conditions and occupant scenarios, and individual thermal images with corresponding TSVs. These datasets include annotated images from multiple experiments, covering simple and complex occupancy scenarios, as well as challenging cases such as overlapping occupants and thermal residual heat imprints. They provide an essential resource for training and evaluating vision-based deep learning models, facilitating further advancements in building management

A deep learning-based framework was developed to conduct vision-based occupancy prediction in indoor spaces. The performance of Shot MultiBox Detector (SSD), Faster Region-based Convolutional Neural Networks (Faster R-CNN), and different versions of You Only Look Once (YOLO) were tested with real-world datasets, demonstrating their strengths and limitations in terms of accuracy, inference time, and generalization

across different building environments. This research contributes a comprehensive evaluation of deep learning algorithms, identifying the optimal trade-offs between model complexity and performance for real-time applications in smart buildings.

A comparative study of vision-based sensors was conducted, systematically analysing the performance of standard and thermal cameras for occupancy prediction. The findings highlight the advantages and limitations of each sensor type, demonstrating that while standard cameras provide higher accuracy and resolution, they suffer from privacy concerns and visual distractions, whereas thermal cameras offer privacy protection and strong performance in low-light conditions but face challenges with overlapping occupants and thermal residual heat imprints. This research establishes a benchmark for vision-based sensor selection, providing a framework for integrating these technologies into building energy management systems.

A new approach to thermal comfort prediction was proposed, with thermal imaging and deep learning to estimate occupant comfort levels without relying on traditional PMV models. By analysing thermal images instead of direct temperature values, the model demonstrated an accuracy of 68.49% in personalized thermal comfort prediction, outperforming PMV-based methods in capturing individual thermal preferences. This approach presents a novel, non-intrusive alternative for real-time occupant comfort assessment, with potential applications in adaptive HVAC control and personalized thermal comfort optimization.

Compared to existing research in the field, this thesis offers several notable advancements in both methodology and application. Many previous studies on occupancy detection have been conducted in highly controlled laboratory environments using static images or manually annotated datasets. In contrast, this research utilised real-time video data collected from multiple field experiments across diverse, naturally

occupied spaces. This allowed for the development and validation of models under realistic conditions, enhancing the ecological validity of the findings.

Furthermore, while most prior work focused on either RGB or thermal cameras in isolation, this thesis presented a systematic comparison between the two modalities. The results provide clear guidance on their relative strengths—highlighting the superior resolution and detection capability of RGB cameras in well-lit conditions, as well as the advantages of thermal cameras in privacy-sensitive and low-light settings.

In the domain of thermal comfort, conventional models such as PMV have been widely adopted, often generalising thermal perception across populations without accounting for personal variability. Recent developments have begun to explore personalised comfort modelling using physiological sensors; however, these typically rely on wearable devices that limit practical deployment. This thesis moves beyond these approaches by introducing a vision-based, non-intrusive method using thermal imaging and deep learning, enabling real-time prediction of individual thermal comfort responses without physical contact or user intervention.

Importantly, while many existing studies remain theoretical or proof-of-concept, this thesis demonstrated the integration of occupancy detection outputs into a simulation environment (IESVE), enabling the assessment of energy consumption and indoor environmental quality based on modelled occupancy profiles. This bridges the gap between machine learning research and practical building management applications—an area that remains underdeveloped in the literature.

Finally, the thesis addresses the emerging concept of dynamic thermal comfort by modelling personalised, time-sensitive comfort responses. While a growing number of studies are beginning to explore adaptive comfort, few have applied deep learning to thermographic data in this way.

This thesis makes a comprehensive contribution to the advancement of intelligent building management by integrating vision-based deep learning methods with occupancy detection, thermal comfort prediction, and energy simulation. It establishes a framework that connects real-time visual sensing, individual comfort modelling, and operational HVAC performance through a sequence of methodologically rigorous experiments. The research introduces practical solutions to limitations in existing approaches—namely, the reliance on static occupancy assumptions, generalised thermal comfort models, and intrusive sensing techniques. By developing and validating non-contact, image-based models for both occupancy and thermal comfort, and demonstrating their application within a building simulation environment, the work moves beyond isolated model development to system-level integration. It contributes novel datasets, experimental workflows, and evaluation metrics that support both academic and industry applications. Collectively, the findings provide a foundation for occupant-aware, adaptive control systems that balance energy efficiency, occupant comfort, and privacy—offering a meaningful step towards the realisation of next-generation smart buildings.

7.3 Overall Study Limitations

This section outlines the key limitations related to data, experimental design, model generalisability, real-world application, and system integration.

Dataset size and diversity present a primary limitation. Although the research involved the creation of several new datasets for occupancy detection and thermal comfort prediction, the overall scale remains not enough compared to large-scale benchmarks in the computer vision domain. The experiments were conducted across a limited number of indoor environments—mainly within the University of Nottingham's campus—restricting the environmental, architectural, and demographic diversity represented in

the data. As a result, the trained models may not generalise well to other building types (e.g., open-plan offices, residential dwellings) or to users with different physical or behavioural profiles. This limitation is particularly relevant for the thermal comfort modelling component, where individual variation plays a significant role.

Thermal comfort experiments were limited to single-occupant settings, chosen to simplify the initial model development and allow for clearer analysis of individual thermal responses. However, shared indoor environments—such as classrooms, offices, or waiting areas—often involve multiple occupants with different comfort preferences. The current experimental design does not account for the thermal interaction effects or the need for group-based comfort balancing strategies. Consequently, the findings may not directly translate to multi-occupant control scenarios without further development.

Another notable limitation lies in cross-subject model performance. While intra-subject results for thermal comfort prediction were promising, cross-subject generalisability proved challenging. Substantial variability was observed in prediction accuracy across different individuals, ranging from over 80% to below 45%. This confirms prior literature findings that thermal sensation is highly subjective and influenced by a range of personal factors, including metabolism, clothing, age, and health status. The current model, although non-intrusive and real-time, still requires further refinement to accommodate broader population diversity.

From a practical implementation perspective, the research did not include real-time deployment or continuous monitoring over extended periods. All video recordings and environmental data were collected in discrete sessions, and model inference was conducted offline. This limits the assessment of model fitness under operational variability such as changes in lighting, occupant behaviour, seasonal temperature shifts, or hardware degradation. Future work should aim to implement and evaluate these models in live settings to test their stability and responsiveness in real-time applications.

Additionally, while the integration of occupancy profiles into IESVE simulations demonstrated the potential for more accurate HVAC control strategies, the system was not connected to an active HVAC unit for closed-loop feedback. Thus, the energy impact assessments remain theoretical, and real-world performance—including HVAC response time, control accuracy, and user feedback—has not been evaluated.

Finally, ethical and privacy considerations were partially addressed through the use of thermal cameras, which avoid facial identification and visible imagery. However, broader concerns around data governance, consent, and user trust were not extensively studied. As vision-based systems become more prevalent in smart buildings, it will be essential to develop guidelines for ethically responsible deployment that respect occupant privacy while still delivering meaningful energy and comfort benefits.

Despite these limitations, the findings of this thesis establish a solid foundation for future work in intelligent, vision-based building systems. They demonstrate clear technical feasibility and uncover important challenges that must be addressed to ensure generalisability, robustness, and practical integration into real-world environments.

7.4 Recommendations for Future Work

Building behaviours and results of this thesis, several areas for future research emerge, particularly in dataset diversity, model generalization across different environments and occupant behaviours, and addressing privacy concerns in vision-based occupancy detection. A focus of future work should be on increasing the size and diversity of training datasets. The experiments showed that dataset expansion played a critical role in improving model performance, particularly in addressing issues such as overlapping occupants and residual heat misclassification. However, the datasets used in this study were limited in terms of environmental diversity, occupant demographics, and activity variations. Future research should aim to collect larger and more diverse datasets across different building types, occupancy densities, and lighting conditions.

While deep learning models performed well in controlled settings, their effectiveness when applied to unseen environments and individuals. To improve adaptability, future research should explore advanced training techniques such as domain adaptation, semi-supervised learning, and synthetic data augmentation. Leveraging transfer learning approaches with pre-trained models could also help reduce the need for extensive labelled datasets while improving generalization to new scenarios.

Privacy concerns remain a major barrier to deploying vision-based occupancy detection in real-world applications. Standard cameras, while effective in capturing high-resolution images, raise ethical and legal concerns regarding personal data collection. This study demonstrated that thermal cameras offer a promising privacy-preserving alternative, as they do not capture identifiable facial features. Future research should refine thermal-based models to further enhance their accuracy and usability. Exploring privacy-aware machine learning techniques, such as federated learning and differential privacy, could also help reduce concerns while ensuring effective real-time occupancy detection.

The application of deep learning models for thermal comfort prediction is another promising direction. While this study demonstrated the feasibility of using thermal imaging for real-time thermal comfort estimation, the dataset used for training was relatively small and focused on a controlled indoor environment, which may not fully capture the variability in individual thermal preferences. Future research should expand the dataset to include a larger and more diverse group of participants, incorporating variations in age, gender, metabolic rate, and clothing insulation. Additionally, investigating differences in thermal comfort across different climate zones could provide valuable insights for developing region-specific comfort prediction models.

Another potential application of thermal comfort modelling is for individuals who have difficulty expressing their comfort preferences, such as young children, the elderly, or

individuals with disabilities. By developing more personalized and adaptive comfort models, future research could explore how thermal imaging, and deep learning can be integrated into assistive technologies to improve occupant well-being. This may require refining models to detect subtle physiological indicators of discomfort, such as changes in skin temperature or posture, and integrating them with intelligent HVAC systems for automated control.

Finally, while this study focused on controlled experimental setups, future research should conduct large-scale field deployments in real-world buildings. Implementing vision-based occupancy detection and thermal comfort models in diverse environments, such as offices, residential spaces, and healthcare facilities, would provide valuable insights into their long-term performance and practicality.

REFERENCES

- C920 HD Pro Webcam* [Online]. [Accessed 2024].
- FLIR ONE® Pro* [Online]. Available: <https://www.flir.co.uk/products/flir-one-pro/?vertical=condition%20monitoring&segment=solutions> [Accessed 2024].
- ABUSHAKRA, B., HABERL, J. S. & CLARIDGE, D. E. J. A. T. 2004. Overview of existing literature on diversity factors and schedules for energy and cooling load calculations. 110, 164-176.
- AC08024865, A. 2005. *Ergonomics of the thermal environment-Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria*, ISO.
- ACQUAAH, Y. T., GOKARAJU, B., TESIERO III, R. C. & MONTY, G. H. 2021. Thermal imagery feature extraction techniques and the effects on machine learning models for smart HVAC efficiency in building energy. *Remote Sensing*, 13, 3847.
- AFTAB, M., CHEN, C., CHAU, C.-K. & RAHWAN, T. 2017. Automatic HVAC Control with Real-time Occupancy Recognition and Simulation-guided Model Predictive Control in Low-cost Embedded System. *Energy and Buildings*, 154.
- AGRAWAL, A., JADHAV, N., GAUR, A., JESWANI, S. & KSHIRSAGAR, A. Improving the Accuracy of Object Detection in Low Light Conditions using Multiple Retinex Theory-based Image Enhancement Algorithms. 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 21-22 April 2022 2022. 1-5.
- AHMAD, T., CHEN, H., GUO, Y. & WANG, J. 2018. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. *Energy and Buildings*, 165, 301–320.
- ALASKAR, H. & SABA, T. Machine Learning and Deep Learning: A Comparative Review. In: SINGH MER, K. K., SEMWAL, V. B., BIJALWAN, V. & CRESPO, R. G., eds. *Proceedings of Integrated Intelligence Enable Networks and Computing*, 2021// 2021 Singapore. Springer Singapore, 143-150.
- ALIERO, M. S., PASHA, M. F., SMITH, D. T., GHANI, I., ASIF, M., JEONG, S. R. & SAMUEL, M. 2022. Non-intrusive room occupancy prediction performance analysis using different machine learning techniques. *Energies*, 15, 9231.
- ALISHAHI, N., OUF, M. M. & NIK-BAKHT, M. 2022. Using WiFi connection counts and camera-based occupancy counts to estimate and predict building occupancy. *Energy and Buildings*, 257, 111759.
- ALOMAIR, B., CLARK, A., CUELLAR, J. & POOVENDRAN, R. Statistical Framework for Source Anonymity in Sensor Networks. 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, 6-10 Dec. 2010 2010. 1-6.

- AMEL, N., MARHIC, B., DELAHOCHÉ, L. & MASSON, J.-B. 2018. ALOS: Automatic learning of an occupancy schedule based on a new prediction model for a smart heating management system. *Building and Environment*, 142.
- AMMER, K. & RING, F. 2019. *The thermal human body: a practical guide to thermal imaging*, Jenny Stanford Publishing.
- APOSTOLO, G., BERNARDINI, F., MAGALHAES, L. & MUCHALUAT-SAADE, D. 2020. *An Experimental Analysis for Detecting Wi-Fi Network Associations Using Multi-label Learning*.
- APOSTOLO, G., BERNARDINI, F., MAGALHAES, L. & MUCHALUAT-SAADE, D. 2021. A Unified Methodology to Predict Wi-Fi Network Usage in Smart Buildings. *IEEE Access*, PP, 1-1.
- ARIEF-ANG, I., SALIM, F. & HAMILTON, M. 2018a. RUP: Large Room Utilisation Prediction with carbon dioxide sensor. *Pervasive and Mobile Computing*, 46.
- ARIEF-ANG, I., SALIM, F. & HAMILTON, M. 2018b. SD-HOC: Seasonal Decomposition Algorithm for Mining Lagged Time Series.
- ARVIDSSON, S., GULLSTRAND, M., SIRMACEK, B. & RIVEIRO, M. 2021. Sensor Fusion and Convolutional Neural Networks for Indoor Occupancy Prediction Using Multiple Low-Cost Low-Resolution Heat Sensor Data. *Sensors*, 21.
- ARYAL, A. & BECERIK-GERBER, B. 2019. A comparative study of predicting individual thermal sensation and satisfaction using wrist-worn temperature sensor, thermal camera and ambient temperature sensor. *Building and Environment*, 160, 106223.
- ARYAL, A. & BECERIK-GERBER, B. 2020. Thermal comfort modeling when personalized comfort systems are in use: Comparison of sensing and learning methods. *Building and Environment*, 185, 107316.
- ASHOURI, A., NEWSHAM, G., SHI, Z. & GUNAY, B. 2019. *Day-ahead Prediction of Building Occupancy using WiFi Signals*.
- ASHRAE 2019. Standard 62.1-2019 -- Ventilation for Acceptable Indoor Air Quality (ANSI Approved).
- AZAR, E. & MENASSA, C. C. 2012. A comprehensive analysis of the impact of occupancy parameters in energy simulation of office buildings. *Energy and Buildings*, 55, 841-853.
- AZULAY, A. & WEISS, Y. 2018. Why do deep convolutional networks generalize so poorly to small image transformations? *arXiv preprint arXiv:1805.12177*.
- BAEK, J., PARK, D. Y., PARK, H., LE, D. M. & CHANG, S. 2023. Vision-based personal thermal comfort prediction based on half-body thermal distribution. *Building and Environment*, 228, 109877.
- BAKANA, S. R., ZHANG, Y. & TWALA, B. 2024. WildARe-YOLO: A lightweight and efficient wild animal recognition model. *Ecological Informatics*, 80, 102541.
- BARTHELMES, V., BECCHIO, C. & CORGNATI, S. 2016. Occupant behavior lifestyles in a residential nearly zero energy building: Effect on energy use and thermal comfort. *Science and Technology for the Built Environment*, 1-16.

- BISONG, E. & BISONG, E. 2019. Google colaboratory. *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners*, 59-64.
- BOSÁK, L. & PALKO, M. 2014. Energy Saving Building in Mountain Area. *Advanced Materials Research*, 1057, 27-34.
- BURZO, M., ABOUELENIEN, M., VAN ALSTINE, D. & RUSINEK, K. Thermal Discomfort Detection Using Thermal Imaging. ASME 2017 International Mechanical Engineering Congress and Exposition, 2017. V006T08A048.
- CALLEMEIN, T., BEECK, K. V. & GOEDEMÉ, T. Anyone here? Smart Embedded Low-Resolution Omnidirectional Video Sensor to Measure Room Occupancy. 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 16-19 Dec. 2019 2019. 1993-2000.
- CAO, X., DAI, X. & LIU, J. 2016. Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. *Energy and Buildings*, 128, 198-213.
- CARLUCCI, S., BAI, L., DE DEAR, R. & YANG, L. 2018. Review of adaptive thermal comfort models in built environmental regulatory documents. *Building and Environment*, 137, 73-89.
- CEN, E. 2019. 16798-1: 2019 Energy Performance of Buildings—Ventilation for Buildings—Part 1: Indoor Environmental Input Parameters for Design and Assessment of Energy Performance of Buildings Addressing Indoor Air Quality. *Thermal Environment, Lighting and Acous.*
- CHAI, Q., WANG, H., ZHAI, Y. & YANG, L. 2020. machine learning algorithms to predict occupants' thermal comfort in naturally ventilated residential buildings. *Energy and Buildings*, 217, 109937.
- CHAI, T. & DRAXLER, R. R. 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, 7, 1247-1250.
- CHARI, A. & CHRISTODOULOU, S. 2017. Building energy performance prediction using neural networks. *Energy Efficiency*, 10, 1315-1327.
- CHEN, S., REN, Y., FRIEDRICH, D., YU, Z. & YU, J. 2021. Prediction of Office Building Electricity Demand using Artificial Neural Network by Splitting the Time Horizon for Different Occupancy Rates. *Energy and AI*, 5, 100093.
- CHEN, Z., JIANG, C. & XIE, L. 2018. Building occupancy estimation and detection: A review. *Energy and Buildings*, 169, 260-270.
- CHEUNG, C. T., PARKINSON, T., LI, P. & BRAGER, G. 2019a. Analysis of the accuracy on PMV – PPD model using the ASHRAE Global Thermal Comfort Database II. *Building and Environment*, 153.
- CHEUNG, T., SCHIAVON, S., PARKINSON, T., LI, P. & BRAGER, G. 2019b. Analysis of the accuracy on PMV – PPD model using the ASHRAE Global Thermal Comfort Database II. *Building and Environment*, 153, 205-217.
- CHIU, Y.-C., TSAI, C.-Y., RUAN, M.-D., SHEN, G.-Y. & LEE, T.-T. Mobilenet-SSDv2: An improved object detection model for embedded systems. 2020

- International conference on system science and engineering (ICSSE), 2020. IEEE, 1-5.
- CHOI, E. J., YUN, J. Y., CHOI, Y. J., SEO, M. C. & MOON, J. W. 2024. Impact of thermal control by real-time PMV using estimated occupants personal factors of metabolic rate and clothing insulation. *Energy and Buildings*, 307, 113976.
- CHOI, H., LEE, J., YI, Y., NA, H., KANG, K. & KIM, T. 2022. Deep vision-based occupancy counting: Experimental performance evaluation and implementation of ventilation control. *Building and Environment*, 223, 109496.
- CHOI, H., UM, C. Y., KANG, K., KIM, H. & KIM, T. 2021. Application of vision-based occupancy counting method using deep learning and performance analysis. *Energy and Buildings*, 252, 111389.
- CHOI, J.-H. & LOFTNESS, V. 2012. Investigation of human body skin temperatures as a bio-signal to indicate overall thermal sensations. *Building and Environment*, 58, 258-269.
- CHOI, J.-H. & YEOM, D. 2017. Study of data-driven thermal sensation prediction model as a function of local body skin temperatures in a built environment. *Building and Environment*, 121, 130-147.
- CHONG, A., AUGENBROE, G. & YAN, D. 2021. Occupancy data at different spatial resolutions: Building energy performance and model calibration. *Applied Energy*, 286, 116492.
- CIBSE 2021. *Chartered Institution of Building Services Engineers*, CIBSE.
- CIBSE, T. 2008. Energy benchmarks. *The Chartered Institution of Building Services Engineers*.
- CLEVENGER, C., HAYMAKER, J. & JALILI, M. 2014. Demonstrating the Impact of the Occupant on Building Performance. *Journal of Computing in Civil Engineering*, 28, 99-102.
- COAKLEY, D., RAFTERY, P. & KEANE, M. 2014. A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews*, 37, 123–141.
- COSMA, A. C. & SIMHA, R. 2018. Thermal comfort modeling in transient conditions using real-time local body temperature extraction with a thermographic camera. *Building and Environment*, 143, 36-47.
- COSMA, A. C. & SIMHA, R. 2019a. Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions. *Building and Environment*, 148, 372-383.
- COSMA, A. C. & SIMHA, R. 2019b. Using the contrast within a single face heat map to assess personal thermal comfort. *Building and Environment*, 160, 106163.
- D'AMBROSIO ALFANO, F. R., PALELLA, B. I. & RICCIO, G. 2011. The role of measurement accuracy on the thermal environment assessment by means of PMV index. *Building and Environment*, 46, 1361-1369.
- DAS, A., SANGOGBOYE, F., KOLVIG-RAUN, E. & KJÆRGAARD, M. 2019. *HeteroSense: An Occupancy Sensing Framework for Multi-Class Classification for Activity Recognition and Trajectory Detection*.

- DAS, P., SHRUBSOLE, C., JONES, B., HAMILTON, I., CHALABI, Z., DAVIES, M., MAVROGIANNI, A. & TAYLOR, J. 2014. Using probabilistic sampling-based sensitivity analyses for indoor air quality modelling. *Building and Environment*, 78.
- DE BOCK, Y., AUQUILLA, A., NOWE, A. & DUFLOU, J. 2020. Nonparametric user activity modelling and prediction. *User Modeling and User-Adapted Interaction*, 30.
- DEVELOPERS, T. 2022. TensorFlow. *Zenodo*.
- DEY, A., LING, X., ADNAN, S., ZHENG, Y., LANDOWSKI, B., ANDERSON, D., STUART, K. & TOLENTINO, M. 2016. *Namataad: Inferring occupancy from building sensors using machine learning*.
- DIAPER, G. 1990. The Hawthorne Effect: a fresh examination. *Educational Studies - EDUC STUD*, 16, 261-267.
- DING, Y., CHEN, W., WEI, S. & YANG, F. 2021a. An occupancy prediction model for campus buildings based on the diversity of occupancy patterns. *Sustainable Cities and Society*, 64, 102533.
- DING, Y., HAN, S., TIAN, Z., YAO, J., CHEN, W. & ZHANG, Q. 2021b. Review on occupancy detection and prediction in building simulation. *Building Simulation*, 15.
- DONG, Z., BOYI, Q., PENGFEI, L. & ZHOUIJIAN, A. 2021. Comprehensive evaluation and optimization of rural space heating modes in cold areas based on PMV-PPD. *Energy and Buildings*, 246, 111120.
- DRIDI, J., AMAYRI, M. & BOUGUILA, N. 2022. Transfer learning for estimating occupancy and recognizing activities in smart buildings. *Building and Environment*, 217, 109057.
- EINI, R. & ABDELWAHED, S. 2019. *Learning-based Model Predictive Control for Smart Building Thermal Management*.
- ERICKSON, V. & CERPA, A. 2012. *Thermovote: Participatory sensing for efficient building HVAC conditioning*.
- ESRAFILIAN-NAJAFABADI, M. & HAGHIGHAT, F. 2021. Occupancy-based HVAC control systems in buildings: A state-of-the-art review. *Building and Environment*, 197, 107810.
- ESRAFILIAN-NAJAFABADI, M. & HAGHIGHAT, F. 2022. Impact of occupancy prediction models on building HVAC control system performance: Application of machine learning techniques. *Energy and Buildings*, 257, 111808.
- FAN, C., XIAO, F. & ZHAO, Y. 2017. A short-term building cooling load prediction method using deep learning algorithms. *Applied Energy*, 195, 222-233.
- FANG, Z., GUO, Z., CHEN, W., WU, H. & ZHENG, Z. 2022. Experimental Investigation of Indoor Thermal Comfort under Different Heating Conditions in Winter. *Buildings*, 12, 2232.
- FANGER, P. 1970a. Thermal Comfort Analysis and Applications in Environment Engineering.
- FANGER, P. O. 1970b. Thermal comfort. Analysis and applications in environmental engineering.

- FERRANTELLI, A., KUIVJÖGI, H., KURNITSKI, J. & THALFELDT, M. 2020. Office Building Tenants' Electricity Use Model for Building Performance Simulations. *Energies*, 13, 5541.
- FERREIRA, P. M., SILVA, S. M., RUANO, A. E., NÉGRIER, A. T. & CONCEIÇÃO, E. Z. E. Neural network PMV estimation for model-based predictive control of HVAC systems. The 2012 International Joint Conference on Neural Networks (IJCNN), 10-15 June 2012 2012. 1-8.
- FÖLDVÁRY LIČINA, V., CHEUNG, T., ZHANG, H., DE DEAR, R., PARKINSON, T., ARENS, E., CHUN, C., SCHIAVON, S., LUO, M., BRAGER, G., LI, P., KAAM, S., ADEBAMOWO, M. A., ANDAMON, M. M., BABICH, F., BOUDEN, C., BUKOVIANSKA, H., CANDIDO, C., CAO, B., CARLUCCI, S., CHEONG, D. K. W., CHOI, J.-H., COOK, M., CROPPER, P., DEUBLE, M., HEIDARI, S., INDRAGANTI, M., JIN, Q., KIM, H., KIM, J., KONIS, K., SINGH, M. K., KWOK, A., LAMBERTS, R., LOVEDAY, D., LANGEVIN, J., MANU, S., MOOSMANN, C., NICOL, F., OOKA, R., OSELAND, N. A., PAGLIANO, L., PETRÁŠ, D., RAWAL, R., ROMERO, R., RIJAL, H. B., SEKHAR, C., SCHWEIKER, M., TARTARINI, F., TANABE, S.-I., THAM, K. W., TELI, D., TOFTUM, J., TOLEDO, L., TSUZUKI, K., DE VECCHI, R., WAGNER, A., WANG, Z., WALLBAUM, H., WEBB, L., YANG, L., ZHU, Y., ZHAI, Y., ZHANG, Y. & ZHOU, X. 2018. Development of the ASHRAE Global Thermal Comfort Database II. *Building and Environment*, 142, 502-512.
- FOUCQUIER, A., ROBERT, S., SUARD, F., STEPHAN, L. & JAY, A. 2013. State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, 23, 272-288.
- FRANCO, A. & LECCESE, F. 2020. Measurement of CO₂ concentration for occupancy estimation in educational buildings with energy efficiency purposes. *Journal of Building Engineering*, 32, 101714.
- FU, H., BALTAZAR, J. & CLARIDGE, D. 2021. Review of developments in whole-building statistical energy consumption models for commercial buildings. *Renewable and Sustainable Energy Reviews*, 147, 111248.
- FUMO, N., MAGO, P. & LUCK, R. 2010. Methodology to estimate building energy consumption using EnergyPlus Benchmark Models. *Energy and Buildings*, 42, 2331-2337.
- GADE, R. & MOESLUND, T. B. 2014. Thermal cameras and applications: a survey. *Machine Vision and Applications*, 25, 245-262.
- GAO, G., GAO, J., LIU, Q., WANG, Q. & WANG, Y. 2020a. Cnn-based density estimation and crowd counting: A survey. *arXiv preprint arXiv:2003.12783*.
- GAO, G., LI, J. & WEN, Y. 2020b. DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning. *IEEE Internet of Things Journal*, 7, 8472-8484.
- GAO, J., ZUO, F., OZBAY, K., HAMMAMI, O. & BARLAS, M. L. 2022. A new curb lane monitoring and illegal parking impact estimation approach based on queueing theory and computer vision for cameras with low resolution and low frame rate. *Transportation Research Part A: Policy and Practice*, 162, 137-154.

- GHAHRAMANI, A., CASTRO, G., KARVIGH, S. A. & BECERIK-GERBER, B. 2018. Towards unsupervised learning of thermal comfort using infrared thermography. *Applied Energy*, 211, 41-49.
- GILANI, S. & GUNAY, B. 2018. Simulating occupants' impact on building energy performance at different spatial scales. *Building and Environment*, 132, 327-337.
- GLAVAŠ, H. Importance of Blackbody in Everyday Infrared Thermography. In: KESER, T., ADEMOVIĆ, N., DESNICA, E. & GRGIĆ, I., eds. 32nd International Conference on Organization and Technology of Maintenance (OTO 2023), 2024// 2024 Cham. Springer Nature Switzerland, 364-374.
- GOYAL, S., INGLE, H. A. & BAROOAH, P. 2012. *Zone-level control algorithms based on occupancy information for energy efficient buildings*.
- GURSEL DINO, I., KALFAOGLU, E., ISERI, O. K., ERDOGAN, B., KALKAN, S. & ALATAN, A. A. 2022. Vision-based estimation of the number of occupants using video cameras. *Advanced Engineering Informatics*, 53, 101662.
- HAIDAR, N., TAMANI, N., NIENABER, F., WESSELING, M. & BOUJU, A. 2019. *Data Collection Period and Sensor Selection Method for Smart Building Occupancy Prediction*.
- HAN, L., FENG, H., LIU, G., ZHANG, A. & HAN, T. 2024. A real-time intelligent monitoring method for indoor evacuee distribution based on deep learning and spatial division. *Journal of Building Engineering*, 92, 109764.
- HE, Y., ZHANG, H., ARENS, E., MERRITT, A., HUIZENGA, C., LEVINSON, R., WANG, A., GHAHRAMANI, A. & ALVAREZ-SUAREZ, A. 2023. Smart detection of indoor occupant thermal state via infrared thermography, computer vision, and machine learning. *Building and Environment*, 228, 109811.
- HITIMANA, E., BAJPAI, G., MUSABE, R., SIBOMANA, L. & JAYAVEL, K. 2021. Implementation of IoT Framework with Data Analysis Using Deep Learning Methods for Occupancy Prediction in a Building. *Future Internet*, 13, 67.
- HONG, T., WANG, Z., LUO, X. & ZHANG, W. 2020. State-of-the-Art on Research and Applications of Machine Learning in the Building Life Cycle. *Energy and Buildings*, 212.
- HOU, H., PAWLAK, J., SIVAKUMAR, A., HOWARD, B. & POLAK, J. 2020. An approach for building occupancy modelling considering the urban context. *Building and Environment*, 183, 107126.
- HSU, T. Y., PHAM, Q. V., CHAO, W. C. & YANG, Y. S. 2020. Post-earthquake building safety evaluation using consumer-grade surveillance cameras. *Smart Structures and Systems, An International Journal*, 25, 531-541.
- HU, S., WANG, P., HOARE, C. & O'DONNELL, J. 2023. Building Occupancy Detection and Localization Using CCTV Camera and Deep Learning. *IEEE Internet of Things Journal*, 10, 597-608.
- HUCHUK, B. & SANNER, S. 2019. Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data. *Building and Environment*, 160, 106177.
- ISSARAVIRIYAKUL, A., PORA, W. & PANITANTUM, N. 2021. *Cloud-based Machine Learning Framework for Residential HVAC Control System*.

- JAZIZADEH, F., GHAHRAMANI, A., BECERIK-GERBER, B., KICHKAYLO, T. & OROSZ, M. 2014. User-led decentralized thermal comfort driven HVAC operations for improved efficiency in office buildings. *Energy and Buildings*, 70, 398-410.
- JAZIZADEH, F. & JUNG, W. 2018. Personalized thermal comfort inference using RGB video images for distributed HVAC control. *Applied Energy*, 220, 829-841.
- JEOUNG, J., JUNG, S., HONG, T., LEE, M. & KOO, C. 2023. Thermal comfort prediction based on automated extraction of skin temperature of face component on thermal image. *Energy and Buildings*, 298, 113495.
- JIANG, L., WANG, X., WANG, L., SHAO, M. & ZHUANG, L. A hybrid ANN-LSTM based model for indoor temperature prediction. 2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA), 1-4 Aug. 2021 2021. 1724-1728.
- JIN, Y., YAN, D., ZHANG, X., AN, J. & HAN, M. 2021. A Data-Driven Model Predictive Control for Lighting System Based on Historical Occupancy in an Office Building: Methodology Development.
- JOCHER, G., CHAURASIA, A., STOKEN, A., BOROVEC, J. & KWON, Y. 2022. ultralytics/yolov5: V6. 1-TensorRT TensorFlow edge TPU and OpenVINO export and inference. *Zenodo*, 2, 2.
- JOCHER, G., CHAURASIA, A., & QIU, J 2023. YOLO by Ultralytics. 8.0.0 ed.
- JOVANOVIĆ, R. Ž., SRETENOVIĆ, A. A. & ŽIVKOVIĆ, B. D. 2015. Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings*, 94, 189-199.
- JU, R.-Y. & CAI, W. 2023. Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm. *Scientific Reports*, 13, 20077.
- JUNG, W. & JAZIZADEH, F. 2020. Energy saving potentials of integrating personal thermal comfort models for control of building systems: Comprehensive quantification through combinatorial consideration of influential parameters. *Applied Energy*, 268, 114882.
- KALLIO, J., TERVONEN, J., RäsÄNEN, P., MÄKYNEN, R., KOIVUSAARI, J. & PELTOLA, J. 2021. Forecasting office indoor CO2 concentration using machine learning with a one-year dataset. *Building and Environment*, 187, 107409.
- KAMEL, E., SHEIKH, S. & HUANG, X. 2020. Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days. *Energy*, 206, 118045.
- KANG, D., MA, Z. & CHAN, A. B. 2019. Beyond Counting: Comparisons of Density Maps for Crowd Analysis Tasks—Counting, Detection, and Tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 29, 1408-1422.
- KARMANN, C., SCHIAVON, S., GRAHAM, L. T., RAFTERY, P. & BAUMAN, F. 2017. Comparing temperature and acoustic satisfaction in 60 radiant and all-air buildings. *Building and Environment*, 126, 431-441.
- KATIĆ, K., LI, R. & ZEILER, W. 2020. Machine learning algorithms applied to a prediction of personal overall thermal comfort using skin temperatures and occupants' heating behavior. *Applied Ergonomics*, 85, 103078.

- KHALIL, M., MCGOUGH, S., POURMIRZA, Z., PAZHOOHESH, M. & WALKER, S. 2021. *Transfer Learning Approach for Occupancy Prediction in Smart Buildings*.
- KILLIAN, M. & KOZEK, M. 2019. Short-term occupancy prediction and occupancy based constraints for MPC of smart homes. *IFAC-PapersOnLine*, 52, 377-382.
- KILLIAN, M., ZAUNER, M. & KOZEK, M. 2018. Comprehensive smart home energy management system using mixed-integer quadratic-programming. *Applied Energy*, 222, 662-672.
- KIM, H. & HONG, T. 2020. Determining the optimal set-point temperature considering both labor productivity and energy saving in an office building. *Applied Energy*, 276, 115429.
- KIM, H., LAMICHHANE, N., KIM, C. & SHRESTHA, R. 2023. Innovations in Building Diagnostics and Condition Monitoring: A Comprehensive Review of Infrared Thermography Applications. *Buildings*, 13, 2829.
- KIM, J. T., LIM, J. H., CHO, S. H. & YUN, G. Y. 2015. Development of the adaptive PMV model for improving prediction performances. *Energy and Buildings*, 98, 100-105.
- KIM, M., JUN, J. A., KIM, N., SONG, Y. J. & PYO, C. S. Sequence-to-Sequence model for Building Energy Consumption Prediction. 2018 International Conference on Information and Communication Technology Convergence (ICTC), 2018.
- KIM, M., KIM, Y., SUNG, S. & YOO, C. 2009. *Data-driven prediction model of indoor air quality by the preprocessed recurrent neural networks*.
- KIM, M. K., KIM, Y.-S. & SREBRIC, J. 2020. Impact of Correlation of Plug Load Data, Occupancy Rates and Local Weather Conditions on Electricity Consumption in a Building Using Four Back-propagation Neural Network Models. *Sustainable Cities and Society*, 62, 102321.
- KIM, S., KANG, S., RYU, K. & SONG, G. 2019. Real-time occupancy prediction in a large exhibition hall using deep learning approach. *Energy and Buildings*, 199.
- KOKLU, M. & TUTUNCU, K. 2019. Tree based classification methods for occupancy detection. *IOP Conference Series: Materials Science and Engineering*, 675, 012032.
- KRAFT, M., ASZKOWSKI, P., PIECZYŃSKI, D. & FULARZ, M. 2021. Low-Cost Thermal Camera-Based Counting Occupancy Meter Facilitating Energy Saving in Smart Buildings. *Energies*, 14, 4542.
- KUSTER, C., REZGUI, Y. & MOURSHED, M. 2017. Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society*, 35.
- LAN, L., LIAN, Z. & PAN, L. 2010. The effects of air temperature on office workers' well-being, workload and productivity-evaluated with subjective ratings. *Applied Ergonomics*, 42, 29-36.
- LAOUADI, A. 2022. A New General Formulation for the PMV Thermal Comfort Index. *Buildings*, 12, 1572.
- LECUN, Y., BENGIO, Y. & HINTON, G. 2015. Deep learning. *Nature*, 521, 436-444.

- LEE, J., WOO, D.-O., JANG, J., JUNG, HANS, L. & LEIGH, S.-B. 2022. Collection and utilization of indoor environmental quality information using affordable image sensing technology. *Energies*, 15, 921.
- LEE, P., SHIN, E.-J., GURALNIK, V., MEHROTRA, S., VENKATASUBRAMANIAN, N. & SMITH, K. T. 2019a. Exploring Privacy Breaches and Mitigation Strategies of Occupancy Sensors in Smart Buildings. *Proceedings of the 1st ACM International Workshop on Technology Enablers and Innovative Applications for Smart Cities and Communities*. New York, NY, USA: Association for Computing Machinery.
- LEE, S., JUNG, S. & LEE, J. 2019b. Prediction Model Based on an Artificial Neural Network for User-Based Building Energy Consumption in South Korea. *Energies*, 12, 608.
- LEE, S., JUNG, S. & LEE, J. J. E. 2019c. Prediction Model Based on an Artificial Neural Network for User-Based Building Energy Consumption in South Korea. 12.
- LI, D., MENASSA, C., KAMAT, V. & BYON, E. 2020a. HEAT - Human Embodied Autonomous Thermostat. *Building and Environment*, 178, 106879.
- LI, D., MENASSA, C. C. & KAMAT, V. R. 2018a. Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography. *Energy and Buildings*, 176, 246-261.
- LI, N., CALIS, G. & BECERIK-GERBER, B. 2012. Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations. *Automation in Construction*, 24, 89-99.
- LI, W., CHEN, J., LAN, F., ZHENG, X. & ZENG, W. 2022. Numerical projection on occupant thermal comfort via dynamic responses to human thermoregulation. *International Journal of Automotive Technology*, 23, 193-203.
- LI, Y., ZHANG, X. & CHEN, D. CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18-23 June 2018 2018b. 1091-1100.
- LI, Z. & DONG, B. 2017. Short Term Predictions of Occupancy in Commercial Buildings—Performance Analysis for Stochastic Models and Machine Learning Approaches. *Energy and Buildings*, 158.
- LI, Z., FRIEDRICH, D. & HARRISON, G. 2020b. Demand Forecasting for a Mixed-Use Building Using Agent-Schedule Information with a Data-Driven Model. *Energies*, 13, 780.
- LIANG, X., HONG, T., SHEN, G. Q. J. B. & ENVIRONMENT 2016. Occupancy data analytics and prediction: A case study. 102, 179-192.
- LIANG, X., SHIM, J., ANDERTON, O. & SONG, D. 2024. Low-cost data-driven estimation of indoor occupancy based on carbon dioxide (CO₂) concentration: A multi-scenario case study. *Journal of Building Engineering*, 82, 108180.
- LO, J. & NOVOSELAC, A. 2010. Localized air-conditioning with occupancy control in an open office. *Energy and Buildings - ENERG BLDG*, 42, 1120-1128.

- LONG, X., DENG, K., WANG, G., ZHANG, Y., DANG, Q., GAO, Y., SHEN, H., REN, J., HAN, S. & DING, E. 2020. PP-YOLO: An effective and efficient implementation of object detector. *arXiv preprint arXiv:2007.12099*.
- LOY-BENITEZ, J., VILELA, P., LI, Q. & YOO, C. 2019. Sequential prediction of quantitative health risk assessment for the fine particulate matter in an underground facility using deep recurrent neural networks. *Ecotoxicology and Environmental Safety*, 169, 316-324.
- LU, X., FENG, F., PANG, Z., YANG, T. & ZHENG, O. 2020. Extracting typical occupancy schedules from social media (TOSSM) and its integration with building energy modeling. *Building Simulation*, 14.
- LUO, N., WANG, Z., BLUM, D., WEYANDT, C., BOURASSA, N., PIETTE, M. A. & HONG, T. 2022. A three-year dataset supporting research on building energy management and occupancy analytics. *Scientific Data*, 9, 156.
- LYDON, D., LYDON, M., TAYLOR, S., DEL RINCON, J. M., HESTER, D. & BROWNJOHN, J. 2019. Development and field testing of a vision-based displacement system using a low cost wireless action camera. *Mechanical Systems and Signal Processing*, 121, 343-358.
- MADDALENA, L., PETROSINO, A. & RUSSO, F. 2014. People counting by learning their appearance in a multi-view camera environment. *Pattern Recognition Letters*, 36, 125-134.
- MALJKOVIC, D. 2019. Modelling Influential Factors of Consumption in Buildings Connected to District Heating Systems. *Energies*, 12, 586.
- MANDIC, D. & CHAMBERS, J. 2001. Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability.
- MANNA, C., FAY, D., BROWN, K. & WILSON, N. 2013. *Learning Occupancy in Single Person Offices with Mixtures of Multi-lag Markov Chains*.
- MAO, N., HAO, J., HE, T., SONG, M., XU, Y. & DENG, S. 2019. PMV-based dynamic optimization of energy consumption for a residential task/ambient air conditioning system in different climate zones. *Renewable Energy*, 142, 41-54.
- MARCHELINA, J., CHOU, S.-Y., YU, V., DEWABHARATA, A., SUGIARTO, V. & KARIJADI, I. 2019. *Two-Stages Occupancy Number Detection Based on Indoor Environment Attributes By Utilizing Machine Learning Algorithm*.
- MASSANA, J., POUS, C., BURGAS, L., MELENDEZ, J. & COLOMER, J. 2016. Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes. *Energy and Buildings*, 130, 519-531.
- MASSIMO, V., GIRETTI, A., TOLVE, L. & CASALS, M. 2016. Model Predictive Energy Control of Ventilation for Underground Stations. *Energy and Buildings*, 116.
- MELFI, R., ROSENBLUM, B., NORDMAN, B. & CHRISTENSEN, K. 2011. Measuring building occupancy using existing network infrastructure. *2011 International Green Computing Conference and Workshops, Orlando, FL, USA*.
- MENG, Y.-B., LI, T.-Y., LIU, G.-H., XU, S.-J. & JI, T. 2020. Real-time dynamic estimation of occupancy load and an air-conditioning predictive control method based on image information fusion. *Building and Environment*, 173, 106741.

- METWALY, A., PEÑA QUERALTA, J., SARKER, V., NGUYEN GIA, T., NASIR, O. & WESTERLUND, T. 2019a. *Edge Computing with Embedded AI: Thermal Image Analysis for Occupancy Estimation in Intelligent Buildings*.
- METWALY, A., QUERALTA, J. P., SARKER, V. K., GIA, T. N., NASIR, O. & WESTERLUND, T. Edge computing with embedded ai: Thermal image analysis for occupancy estimation in intelligent buildings. *Proceedings of the INTelligent Embedded Systems Architectures and Applications Workshop 2019*, 2019b. 1-6.
- MITRA, A., NGOKO, Y. & TRYSTRAM, D. Smart Oracle Based Building Management System. 2021 IEEE International Conference on Smart Computing (SMARTCOMP), 23-27 Aug. 2021 2021. 61-68.
- MOHRI, M., ROSTAMIZADEH, A. & TALWALKAR, A. 2012. Foundations of Machine Learning.
- MOON, J. W. 2012. Performance of ANN-based predictive and adaptive thermal-control methods for disturbances in and around residential buildings. *Building and Environment*, 48, 15-26.
- MTIBAA, F., NGUYEN, K.-K., AZAM, M., PAPACHRISTOU, A., VENNE, J.-S. & CHERIET, M. 2020. LSTM-based indoor air temperature prediction framework for HVAC systems in smart buildings. *Neural Computing and Applications*, 32.
- MUJAN, I., ANĐELKOVIĆ, A. S., MUNČAN, V., KLJAJIĆ, M. & RUŽIĆ, D. 2019. Influence of indoor environmental quality on human health and productivity - A review. *Journal of Cleaner Production*, 217, 646-657.
- MUMMA, S. A. 2004. Transient occupancy ventilation by monitoring CO₂. *ASHRAE IAQ Applications*, 21-23.
- MURAKAMI, Y., TERANO, M., MIZUTANI, K., HARADA, M. & KUNO, S. 2007. Field experiments on energy consumption and thermal comfort in the office environment controlled by occupants' requirements from PC terminal. *Building and Environment*, 42, 4022-4027.
- NASTASI, B., MARKOVSKA, N., PUKSEC, T., DUIĆ, N. & FOLEY, A. 2022. Renewable and sustainable energy challenges to face for the achievement of Sustainable Development Goals. *Renewable and Sustainable Energy Reviews*, 157, 112071.
- NGUYEN, T. A. & AIELLO, M. 2013. Energy intelligent buildings based on user activity: A survey. *Energy and Buildings*, 56, 244-257.
- NICOL, F. & WILSON, M. An overview of the European Standard EN 15251. WINDSOR CONFERENCE, 2010.
- NKURIKIYEYEU, K. N., SUZUKI, Y., TOBE, Y., LOPEZ, G. F. & ITAO, K. Heart rate variability as an indicator of thermal comfort state. 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 19-22 Sept. 2017 2017. 1510-1512.
- O'MAHONY, N., CAMPBELL, S., CARVALHO, A., HARAPANAHALLI, S., HERNANDEZ, G. V., KRPALKOVA, L., RIORDAN, D. & WALSH, J. Deep Learning vs. Traditional Computer Vision. *In: ARAI, K. & KAPOOR, S., eds.*

- Advances in Computer Vision, 2020// 2020 Cham. Springer International Publishing, 128-144.
- OLESEN, B. W. & BRAGER, G. S. 2004. A better way to predict comfort: The new ASHRAE standard 55-2004.
- OMAROV, B., ALTAYEVA, A., DEMEUOV, A., TASTANOV, A., KASSYMBEKOV, Z. & KOISHYBAYEV, A. Fuzzy Controller for Indoor Air Quality Control: A Sport Complex Case Study. *In*: LUHACH, A. K., JAT, D. S., BIN GHAZALI, K. H., GAO, X.-Z. & LINGRAS, P., eds. Advanced Informatics for Computing Research, 2021// 2021 Singapore. Springer Singapore, 53-61.
- PANCHABIKESEN, K., HAGHIGHAT, F. & MANKIBI, M. 2020. Data driven occupancy information for energy simulation and energy use assessment in residential building. *Energy*, 218.
- PANG, Z., O'NEILL, Z., CHEN, Y., ZHANG, J., CHENG, H. & DONG, B. 2023. Adopting occupancy-based HVAC controls in commercial building energy codes: Analysis of cost-effectiveness and decarbonization potential. *Applied Energy*, 349, 121594.
- PAPPALARDO, M. & REVERDY, T. 2020. Explaining the performance gap in a French energy efficient building: Persistent misalignment between building design, space occupancy and operation practices. *Energy Research & Social Science*, 70, 101809.
- PARK, H. & PARK, D. Y. 2022. Prediction of individual thermal comfort based on ensemble transfer learning method using wearable and environmental sensors. *Building and Environment*, 207, 108492.
- PARK, J. Y., OUF, M. M., GUNAY, B., PENG, Y., O'BRIEN, W., KJÆRGAARD, M. B. & NAGY, Z. 2019. A critical review of field implementations of occupant-centric building controls. *Building and Environment*, 165, 106351.
- PASZKE, A., GROSS, S., MASSA, F., LERER, A., BRADBURY, J., CHANAN, G., KILLEEN, T., LIN, Z., GIMELSHEIN, N. & ANTIGA, L. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- PENG, Y., RYSANEK, A., NAGY, Z. & SCHLUETER, A. 2017. Occupancy learning-based demand-driven cooling control for office spaces. *Building and Environment*, 122.
- PESIC, S., TOSIC, M., IKOVIĆ, O., RADOVANOVIC, M., IVANOVIC, M. & BOSCOVIC, D. 2019. BLEMAT: Data Analytics and Machine Learning for Smart Building Occupancy Detection and Prediction. *International Journal on Artificial Intelligence Tools*, 28, 1960005.
- PIGLIAUTILE, I., CASACCIA, S., MORRESI, N., ARNESANO, M., PISELLO, A. L. & REVEL, G. 2020. Assessing occupants' personal attributes in relation to human perception of environmental comfort: Measurement procedure and data analysis. *Building and Environment*, 177, 106901.

- QIAO, L. & YAN, X. 2022. Analysis of Thermal Comfort under Different Exercise Modes in Winter in Universities in Severe Cold Regions. *Sustainability*, 14, 15796.
- QIN, Z., CHAKI, D., LAKHDARI, A., ABUSAFIA, A. & BOUGUETTAYA, A. Occupancy estimation from thermal images. *International Conference on Service-Oriented Computing*, 2021. Springer, 301-305.
- RAHAMAN, M., PARE, H., LIONO, J., SALIM, F., REN, Y., CHAN, J., KUDO, S., RAWLING, T. & SINICKAS, A. 2019. *OccuSpace: Towards a Robust Occupancy Prediction System for Activity Based Workplace*.
- RAMOKONE, A., POPOOLA, O. & AWELEWA, A. 2020. *An intelligent approach for assessing occupancy and occupant-related activities impact on residential electric load profiles*.
- RAMOKONE, A., POPOOLA, O., AWELEWA, A. & AYODELE, T. 2021. A review on behavioural propensity for building load and energy profile development – Model inadequacy and improved approach. *Sustainable Energy Technologies and Assessments*, 45, 101235.
- RANJAN, J. & SCOTT, J. 2016. ThermalSense: determining dynamic thermal comfort preferences using thermographic imaging. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Heidelberg, Germany: Association for Computing Machinery.
- RAZAVI, R., GHARIPOUR, A., FLEURY, M. & AKPAN, I. J. 2019. Occupancy detection of residential buildings using smart meter data: A large-scale study. *Energy and Buildings*, 183, 195-208.
- RAZBAN, A. & TAHERI, S. 2021. Learning-based CO₂ concentration prediction: Application to indoor air quality control using demand-controlled ventilation. *Building and Environment*, 205.
- REBAÑO-EDWARDS, S. 2007. Modelling perceptions of building quality—a neural network approach. *Building and Environment*, 42(7), 2762-2777. *Building and Environment*, 42, 2762-2777.
- REENA, M., MATHEW, D. & JACOB, L. 2018. A flexible control strategy for energy and comfort aware HVAC in large buildings. *Building and Environment*, 145.
- REN, S., HE, K., GIRSHICK, R. & SUN, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.
- ROBOFLOW, I. 2023. *Roboflow Universe* [Online]. Available: <https://universe.roboflow.com/> [Accessed].
- ROULET, C.-A., FLOURENTZOU, F., FORADINI, F., BLUYSSSEN, P., COX, C. & AIZLEWOOD, C. 2006. Multicriteria analysis of health, comfort and energy efficiency in buildings. *Building Research and Information*, 34.
- RUEDA, L., AGBOSSOU, K., HENAO, N. F., KELOUWANI, S., OVIEDO-CEPEDA, J., LOSTEC, B., SANSREGRET, S. & FOURNIER, M. 2021. Online Unsupervised Occupancy Anticipation System Applied to Residential Heat Load Management. *IEEE Access*, 9, 109806-109821.

- RYU, S. H. & MOON, H. J. 2016. Development of an occupancy prediction model using indoor environmental data based on machine learning techniques. *Building and Environment*, 107, 1-9.
- SAHA, H., FLORITA, A., HENZE, G. & SARKAR, S. 2019. Occupancy Sensing in Buildings: A Review of Data Analytics Approaches. *Energy and Buildings*, 188-189.
- SAHOO, S. R. & LONE, H. R. Occupancy counting in dense and sparse settings with a low-cost thermal camera. 2023 15th International Conference on COMMunication Systems & NETworkS (COMSNETS), 2023. IEEE, 537-544.
- SALEHI, B., GHANBARAN, A. H. & MAEREFAT, M. 2020. Intelligent models to predict the indoor thermal sensation and thermal demand in steady state based on occupants' skin temperature. *Building and Environment*, 169, 106579.
- SALIMI, S., LIU, Z. & HAMMAD, A. 2019. Occupancy prediction model for open-plan offices using real-time location system and inhomogeneous Markov chain. *Building and Environment*, 152.
- SAMA, S. & RAHNAMAY-NAEINI, M. 2016. *A study on compression-based sequential prediction methods for occupancy prediction in smart homes*.
- SANGOGBOYE, F., ARENDT, K., SINGH, A., VEJE, C., KJÆRGAARD, M. & JØRGENSEN, B. 2017. Performance comparison of occupancy count estimation and prediction with common versus dedicated sensors for building model predictive control. *Building Simulation*, 10, 1-15.
- SBCI, U. 2009. Buildings and climate change: summary for decision-makers. *Buildings and Climate Change Summary for Decision-makers*, 1-62.
- SHAN, X. & YANG, E.-H. 2020. Supervised machine learning of thermal comfort under different indoor temperatures using EEG measurements. *Energy and Buildings*, 225, 110305.
- SHEIKH KHAN, D., KOLARIK, J., ANKER HVIID, C. & WEITZMANN, P. 2021. Method for long-term mapping of occupancy patterns in open-plan and single office spaces by using passive-infrared (PIR) sensors mounted below desks. *Energy and Buildings*, 230, 110534.
- ŠIMIĆ, G. & DEVEDŽIĆ, V. 2003. Building an intelligent system using modern Internet technologies. *Expert Systems with Applications*, 25, 231-246.
- SINDAGI, V. A. & PATEL, V. M. 2018. A survey of recent advances in CNN-based single image crowd counting and density estimation. *Pattern Recognition Letters*, 107, 3-16.
- SIRMACEK, B. & RIVEIRO, M. 2020. Occupancy Prediction Using Low-Cost and Low-Resolution Heat Sensors for Smart Offices. *Sensors*, 20, 5497.
- SOKOLOVA, M., JAPKOWICZ, N. & SZPAKOWICZ, S. Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. AI 2006: Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, December 4-8, 2006. Proceedings 19, 2006. Springer, 1015-1021.
- SOLUTIONS, I. E. 2020. IES Virtual Environment (IESVE)[Computer software].

- STANDARD, A. 1992. Thermal environmental conditions for human occupancy. *ANSI/ASHRAE*, 55, 5.
- STATISTICS, U. S. B. O. L. 2009. American Time Use Survey.
- SULTAN, Z., PANTAZARAS, A., CHATURVEDI, K. A., YANG, J., THAM, K. & LEE, S. E. 2017. Predicting occupancy counts using physical and statistical Co₂-based modeling methodologies. *Building and Environment*, 123.
- SUN, K., LIU, P., XING, T., ZHAO, Q. & WANG, X. 2022a. A fusion framework for vision-based indoor occupancy estimation. *Building and Environment*, 225, 109631.
- SUN, K., YAN, D., HONG, T. & GUO, S. 2014. Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration. *Building and Environment*, 79, 1-12.
- SUN, K., ZHAO, Q., ZHANG, Z. & HU, X. 2022b. Indoor occupancy measurement by the fusion of motion detection and static estimation. *Energy and Buildings*, 254, 111593.
- SUN, K., ZHAO, Q. & ZOU, J. 2020. A review of building occupancy measurement systems. *Energy and Buildings*, 216, 109965.
- TAGLIABUE, L., RE CECCONI, F., RINALDI, S. & CIRIBINI, A. 2021. Data driven Indoor air quality prediction in educational facilities based on IoT network. *Energy and Buildings*, 236, 110782.
- TALON, C. & GOLDSTEIN, N. 2015. Smart offices: how intelligent building solutions are changing the occupant experience. *Navigant Consulting*.
- TAMURA, K., OBA, Y., ARX, T. & LOZANOFF, S. 2018. The Face – A Vascular Perspective. *Swiss Dental Journal*, 128, 382-392.
- TARTARINI, F., SCHIAVON, S., CHEUNG, T. & HOYT, T. 2020. CBE Thermal Comfort Tool: Online tool for thermal comfort calculations and visualizations. *SoftwareX*, 12, 100563.
- TIEN, P., CALAUTIT, J. K., DARKWA, J., WOOD, C., WEI, S., PANTUA, C. & XU, W. 2020a. *A deep learning framework for energy management and optimisation of HVAC systems*.
- TIEN, P., WEI, S., CALAUTIT, J. K., DARKWA, J. & WOOD, C. 2020b. 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*, N/A.
- TIEN, P., WEI, S., CALAUTIT, J. K., DARKWA, J. & WOOD, C. 2020c. A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand-driven control solutions. *Energy and Buildings*, 226, 110386.
- TIEN, P. W., WEI, S., CALAUTIT, J. K., DARKWA, J. & WOOD, C. 2020d. A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand-driven control solutions. *Energy and Buildings*, 226, 110386.
- TIEN, P. W., WEI, S., CALAUTIT, J. K., DARKWA, J. & WOOD, C. 2022. Real-time monitoring of occupancy activities and window opening within buildings using

- an integrated deep learning-based approach for reducing energy demand. *Applied Energy*, 308, 118336.
- TIEN, P. W., WEI, S., LIU, T., CALAUTIT, J., DARKWA, J. & WOOD, C. 2021. 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.
- TONG, R., XIE, D. & TANG, M. 2013. Upper Body Human Detection and Segmentation in Low Contrast Video. *IEEE Transactions on Circuits and Systems for Video Technology*, 23, 1502-1509.
- TRUONG, L., CHOW, K., LUEVISADPAIBUL, R., THIRUNAVUKKARASU, G., SEYEDMAHMOUDIAN, M., HORAN, B., MEKHILEF, S. & STOJCEVSKI, A. 2021. Accurate Prediction of Hourly Energy Consumption in a Residential Building Based on the Occupancy Rate Using Machine Learning Approaches. *Applied Sciences*, 11, 2229.
- TUOHY, P., HUMPHREYS, M., NICOL, F., RIJAL, H. & CLARKE, J. A. 2009. Occupant behaviour in naturally ventilated and hybrid buildings.
- TURLEY, C., JACOBY, M., HENZE, G. & PAVLAK, G. 2020. Development and evaluation of occupancy-aware model predictive control for residential building energy efficiency and occupant comfort. *IOP Conference Series: Earth and Environmental Science*, 588, 022043.
- TZUTALIN 2018. LabelImg.
- ULTRALYTICS. 2023. *Ultralytics YOLO Docs* [Online]. Available: <https://docs.ultralytics.com/> [Accessed].
- VAŇUŠ, J., MAJIDZADEH GORJANI, O. & BILIK, P. 2019. Novel Proposal for Prediction of CO₂ Course and Occupancy Recognition in Intelligent Buildings within IoT. *Energies*, 12, 4541.
- WANG, A., CHEN, H., LIU, L., CHEN, K., LIN, Z., HAN, J. & DING, G. 2024. YOLOv10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*.
- WANG, C.-Y., BOCHKOVSKIY, A. & LIAO, H.-Y. M. 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*.
- WANG, W., CHEN, J. & HONG, T. 2018a. Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings. *Automation in Construction*, 94.
- WANG, W., CHEN, J., HONG, T. & ZHU, N. 2018b. Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology. *Building and Environment*, 138.
- WANG, W., CHEN, J. & SONG, X. 2017. Modeling and Predicting Occupancy Profile in Office Space with a Wi-Fi Probe-based Dynamic Markov Time-window Inference Approach. *Building and Environment*, 124.
- WANG, W., HONG, T., LI, N., WANG, R. & CHEN, J. 2018c. Linking energy-cyber-physical systems with occupancy prediction and interpretation through WiFi probe-based ensemble classification. *Applied Energy*, 236.

- WANG, W., HONG, T., XU, N., XU, X., CHEN, J. & SHAN, X. 2019a. Cross-source sensing data fusion for building occupancy prediction with adaptive lasso feature filtering. *Building and Environment*, 162.
- WANG, X. & TAGUE, P. 2014. Non-Invasive User Tracking via Passive Sensing: Privacy Risks of Time-Series Occupancy Measurement. *Proceedings of the 2014 Workshop on Artificial Intelligent and Security Workshop*. Scottsdale, Arizona, USA: Association for Computing Machinery.
- WANG, Y., WANG, Y., LIU, L. & CHEN, X. 2019b. Defocused camera calibration with a conventional periodic target based on Fourier transform. *Optics Letters*, 44, 3254-3257.
- WANG, Z., WANG, J., HE, Y., LIU, Y., LIN, B. & HONG, T. 2020. Dimension analysis of subjective thermal comfort metrics based on ASHRAE Global Thermal Comfort Database using machine learning. *Journal of Building Engineering*, 29, 101120.
- WANG, Z., WANG, Y. & SRINIVASAN, R. S. 2018d. A novel ensemble learning approach to support building energy use prediction. *Energy and Buildings*, 159, 109-122.
- WEI, S., TIEN, P. W., CHOW, T. W., WU, Y. & CALAUTIT, J. K. 2022a. Deep learning and computer vision based occupancy CO₂ level prediction for demand-controlled ventilation (DCV). *Journal of Building Engineering*, 56, 104715.
- WEI, S., TIEN, P. W., WU, Y. & CALAUTIT, J. K. 2022b. A coupled deep learning-based internal heat gains detection and prediction method for energy-efficient office building operation. *Journal of Building Engineering*, 47, 103778.
- WEI, W., RAMALHO, O., MALINGRE, L., SIVANANTHAM, S., LITTLE, J. & MANDIN, C. 2019a. Machine learning and statistical models for predicting indoor air quality. *Indoor Air*, 29.
- WEI, Y., XIA, L., PAN, S., WU, J., ZHANG, X., HAN, M., ZHANG, W., XIE, J. & LI, Q. 2019b. Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks. *Applied Energy*, 240, 276-294.
- WEI, Y., ZHANG, X., SHI, Y., XIA, L., PAN, S., WU, J., HAN, M. & ZHAO, X. 2018. A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, 82, 1027-1047.
- WINKLER, D., YADAV, A., CHITU, C. & CERPA, A. 2020. *OFFICE: Optimization Framework For Improved Comfort & Efficiency*.
- WINKLER, T. & RINNER, B. 2013. Privacy and security in video surveillance. *Intelligent Multimedia Surveillance: Current Trends and Research*, 37-66.
- WU, J., WEI, Y. & ZHANG, X. 2021. Prediction of Occupancy Level and Energy Consumption in Office Building Using Blind System Identification and Neural Networks.

- WU, Y., CAO, B., HU, M., LV, G., MENG, J. & ZHANG, H. 2023a. Development of personal comfort model and its use in the control of air conditioner. *Energy and Buildings*, 285, 112900.
- WU, Y., ZHANG, Z., LIU, H., LI, B., CHEN, B., KOSONEN, R. & JOKISALO, J. 2023b. Age differences in thermal comfort and physiological responses in thermal environments with temperature ramp. *Building and Environment*, 228, 109887.
- WU, Z., LI, N., PENG, J., CUI, H., LIU, P., LI, H. & LI, X. 2018. Using an ensemble machine learning methodology-Bagging to predict occupants' thermal comfort in buildings. *Energy and Buildings*, 173, 117-127.
- WUXIA 2022. People_small Dataset. *Roboflow Universe*.
- XIE, J., LI, H., LI, C., ZHANG, J. & LUO, M. 2020. Review on occupant-centric thermal comfort sensing, predicting, and controlling. *Energy and Buildings*, 226, 110392.
- XILEI, D., LIU, J. & ZHANG, X. 2020. A review of studies applying machine learning models to predict occupancy and window-opening behaviours in smart buildings. *Energy and Buildings*, 223, 110159.
- XIONG, J., LIAN, Z., ZHOU, X., YOU, J. & LIN, Y. 2016. Potential indicators for the effect of temperature steps on human health and thermal comfort. *Energy and Buildings*, 113, 87-98.
- YAMAZAKI, F., INOUE, K., OHMI, N. & OKIMOTO, C. 2023. A two-week exercise intervention improves cold symptoms and sleep condition in cold-sensitive women. *Journal of Physiological Anthropology*, 42, 22.
- YANG, L., YE, M. & HE, B.-J. 2014. CFD simulation research on residential indoor air quality. *Science of The Total Environment*, 472, 1137-1144.
- YANG, Y., YUAN, Y., PAN, T., ZANG, X. & LIU, G. 2022. A framework for occupancy prediction based on image information fusion and machine learning. *Building and Environment*, 207, 108524.
- YANG, Z. & BECERIK-GERBER, B. 2014. The coupled effects of personalized occupancy profile based HVAC schedules and room reassignment on building energy use. *Energy and Buildings*, 78, 113–122.
- YAO, R., LI, B. & LIU, J. 2009. A theoretical adaptive model of thermal comfort – Adaptive Predicted Mean Vote (aPMV). *Building and Environment*, 44, 2089-2096.
- YAO, Y., LIAN, Z., LIU, W. & SHEN, Q. 2008. Experimental study on physiological responses and thermal comfort under various ambient temperatures. *Physiology & Behavior*, 93, 310-321.
- YAO, Y. & SHEKHAR, D. 2021. State of the art review on model predictive control (MPC) in Heating Ventilation and Air-conditioning (HVAC) field. *Building and Environment*, 200, 107952.
- YAU, Y. H. & CHEW, B. T. 2012. A review on predicted mean vote and adaptive thermal comfort models. *Building Services Engineering Research and Technology*, 35, 23-35.

- YOSHINO, H., HONG, T. & NORD, N. 2017. IEA EBC Annex 53: Total Energy Use in Buildings – Analysis and Evaluation Methods. *Energy and Buildings*, 152.
- YU, Z., FUNG, B., HAGHIGHAT, F., YOSHINO, H. & MOROFSKY, E. 2011. A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and Buildings*, 43, 1409-1417.
- YUAN, Y., SHIM, J., LEE, S., SONG, D. & KIM, J. 2020. Prediction for Overheating Risk Based on Deep Learning in a Zero Energy Building. *Sustainability*, 12, 8974.
- ZANTALIS, F., KOULOURLAS, G., KARABETSOS, S. & KANDRIS, D. 2019. A Review of Machine Learning and IoT in Smart Transportation. *Future Internet*, 11, 94.
- ZENG, A., LIU, S. & YU, Y. 2019. Comparative Study of Data Driven Methods in Building Electricity Use Prediction. *Energy and Buildings*, 194.
- ZHANG, F., HADDAD, S., NAKISA, B., RASTGOO, M. N., CANDIDO, C., TJONDRONEGORO, D. & DE DEAR, R. 2017. The effects of higher temperature setpoints during summer on office workers' cognitive load and thermal comfort. *Building and Environment*, 123, 176-188.
- ZHANG, W., CALAUTIT, J., TIEN, P. W., WU, Y. & WEI, S. 2024. Deep Learning Models for Vision-Based Occupancy Detection in High Occupancy Buildings. *Journal of Building Engineering*, 111355.
- ZHANG, W., HU, W. & WEN, Y. 2019. Thermal Comfort Modeling for Smart Buildings: A Fine-Grained Deep Learning Approach. *IEEE Internet of Things Journal*, 6, 2540-2549.
- ZHANG, W., WU, Y. & CALAUTIT, J. K. 2022. A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment. *Renewable and Sustainable Energy Reviews*, 167, 112704.
- ZHANG, Y., ZHOU, D., CHEN, S., GAO, S. & MA, Y. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June 2016 2016. 589-597.
- ZHAO, H.-X. & MAGOULÈS, F. 2012. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16, 3586-3592.
- ZHOU, X., XU, L., ZHANG, J., NIU, B., LUO, M., ZHOU, G. & ZHANG, X. 2020. Data-driven thermal comfort model via support vector machine algorithms: Insights from ASHRAE RP-884 database. *Energy and Buildings*, 211, 109795.
- ZOU, J., ZHAO, Q., YANG, W. & WANG, F. 2017. Occupancy detection in the office by analyzing surveillance videos and its application to building energy conservation. *Energy and Buildings*, 152, 385-398.

APPENDICES

Appendix.A .

This appendix presents the training loss curve results for Faster R-CNN, SSD, YOLOv5n, YOLOv5x, YOLOv7, YOLOv7w6, YOLOv8n, and YOLOv8x in Chapter 3.

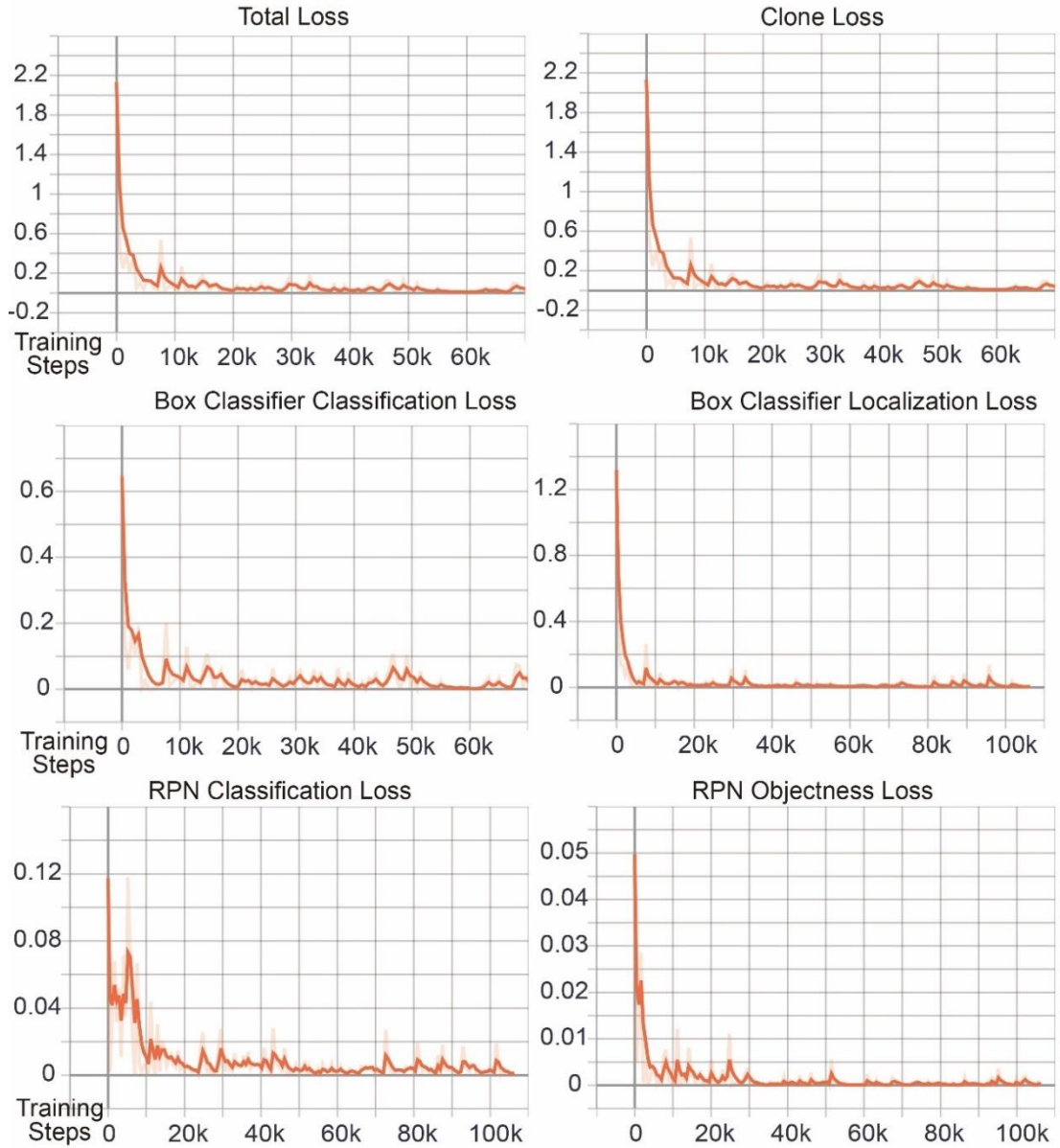


Figure A-1 The FasterRCNN result metrics for the training sets: Total loss, Clone loss, Box classifier classification loss, Box classifier localization loss, RPN classification loss, and RPN objectless loss.

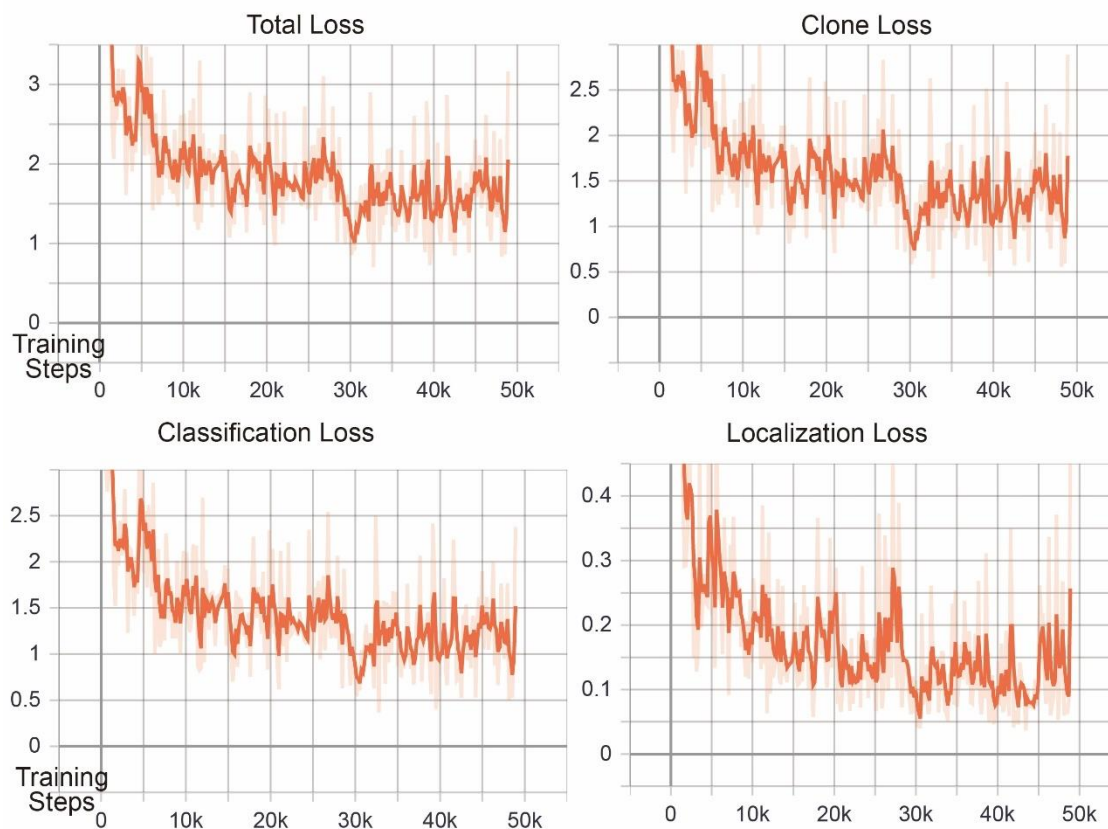


Figure A-2 The SSD result metrics for the training sets: Total loss, Clone loss, Classification loss and Localization loss.

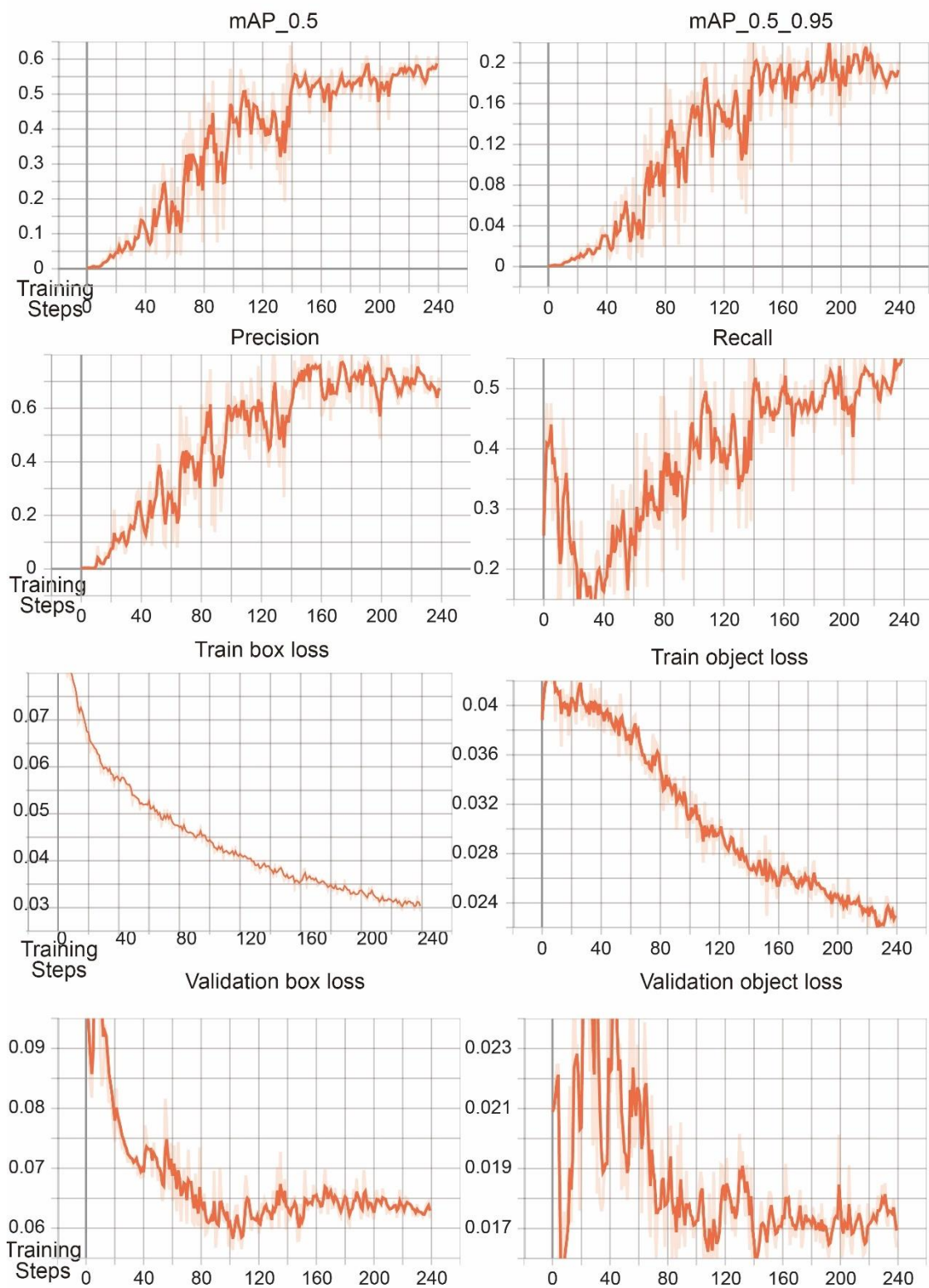


Figure A-3 The YOLOv5n result metrics for the training and validation sets: box loss, classification loss, objectness loss, precision, recall and mAP.

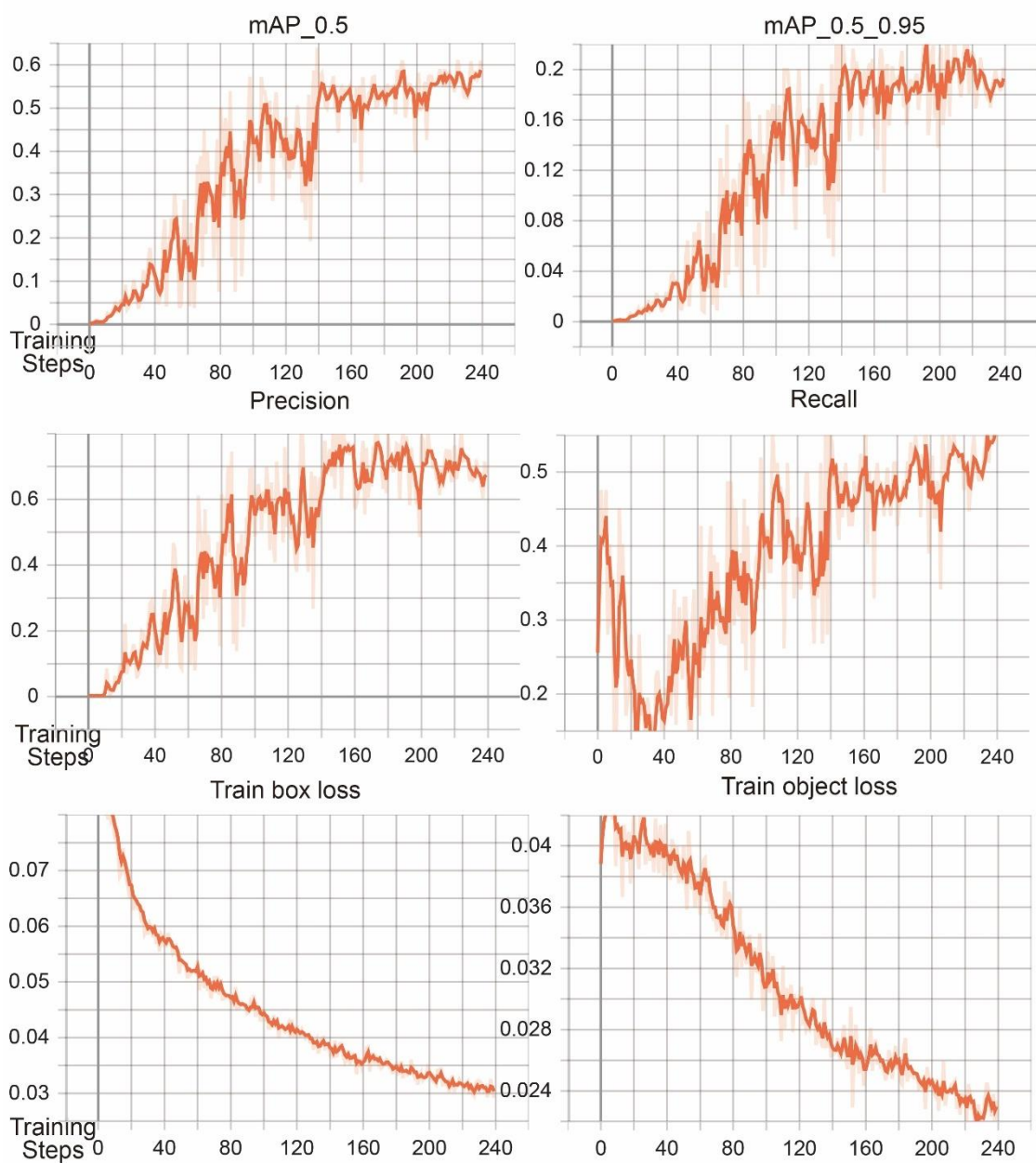


Figure A-4 The YOLOv5x result metrics for the training sets: box loss, classification loss, objectness loss, precision, recall and mAP.

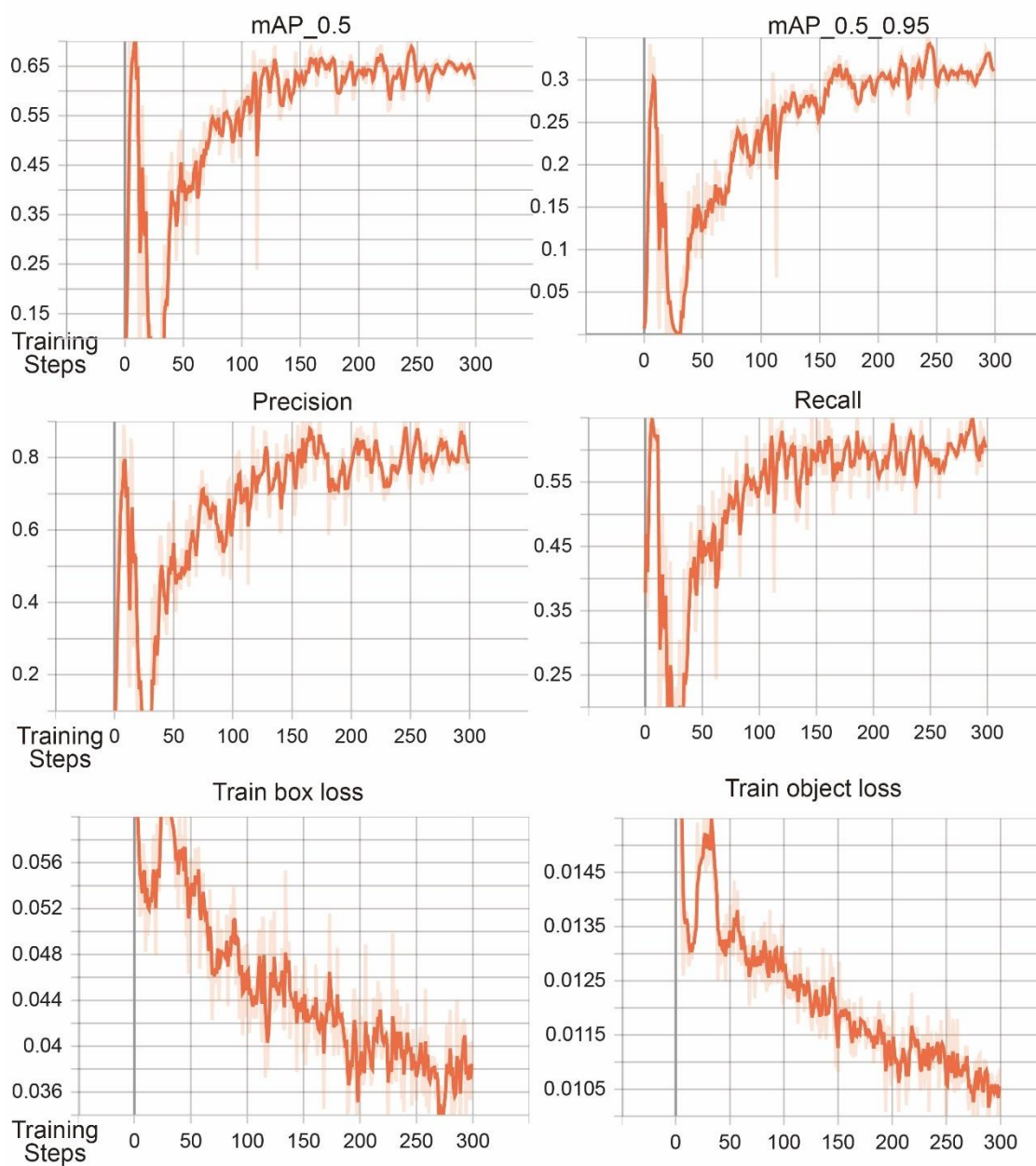


Figure A-5 The YOLOv7 result metrics for the training sets: box loss, classification loss, objectness loss, precision, recall and mAP.

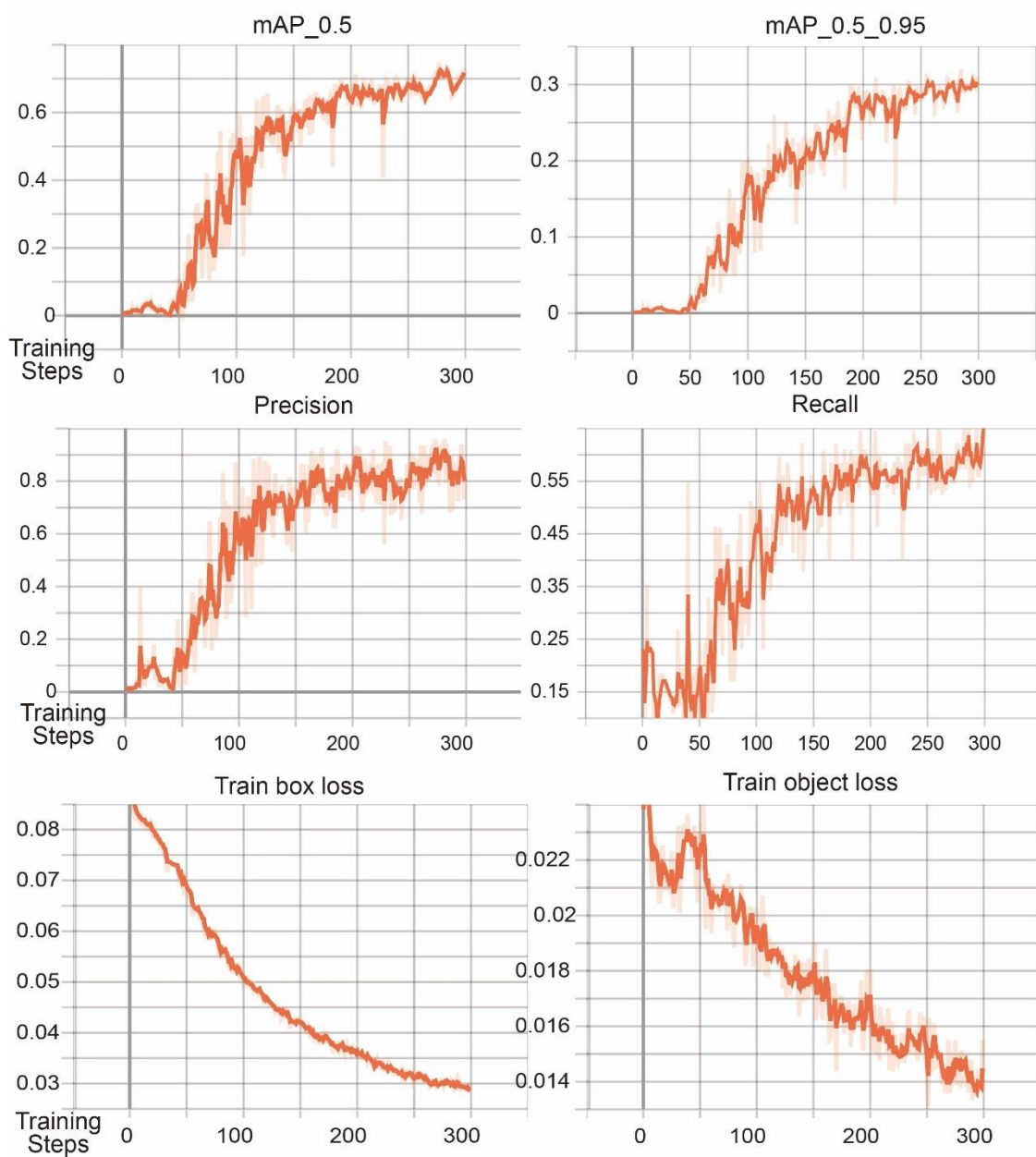


Figure A-6 The YOLOv7w6 result metrics for the training sets: box loss, classification loss, objectness loss, precision, recall and mAP.

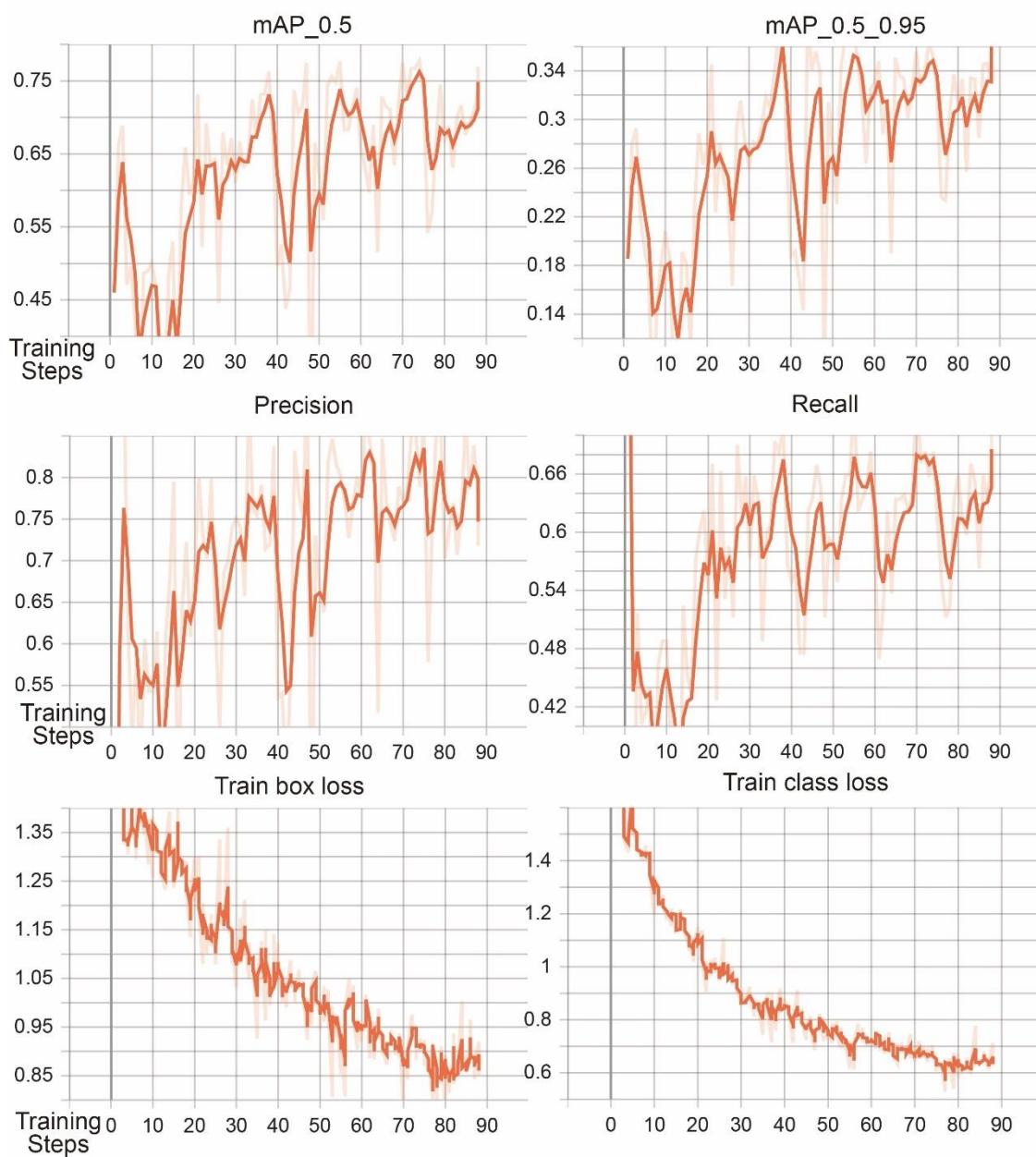


Figure A-7 The YOLOv8n result metrics for the training sets: box loss, classification loss, objectness loss, precision, recall and mAP.

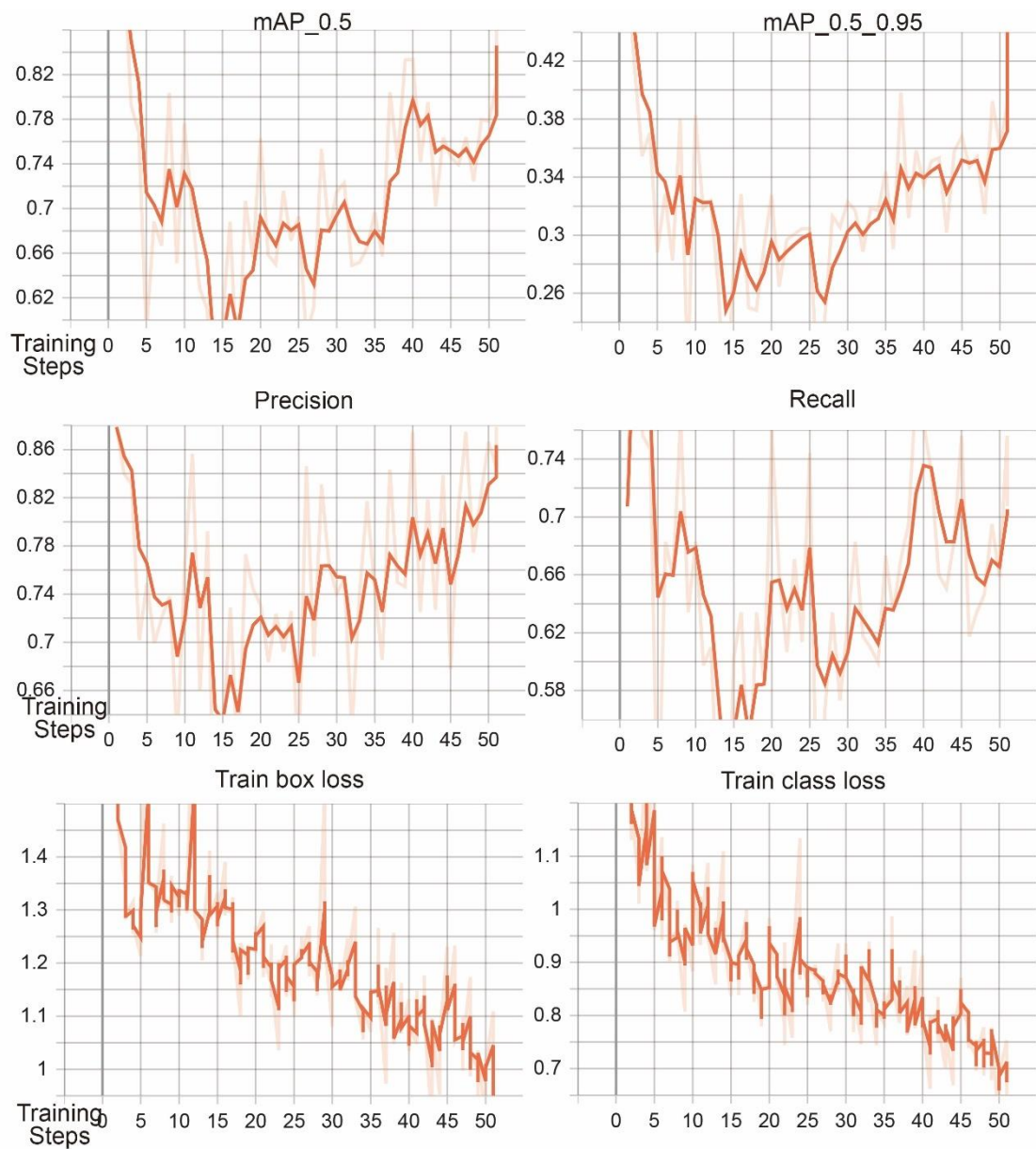


Figure A-8 The YOLOv8x result metrics for the training sets: box loss, classification loss, objectness loss, precision, recall and mAP.

Appendix.B .

This appendix presents the normalized confusion matrix in training and validation for 14 intra-subject experiments in Chapter 5.

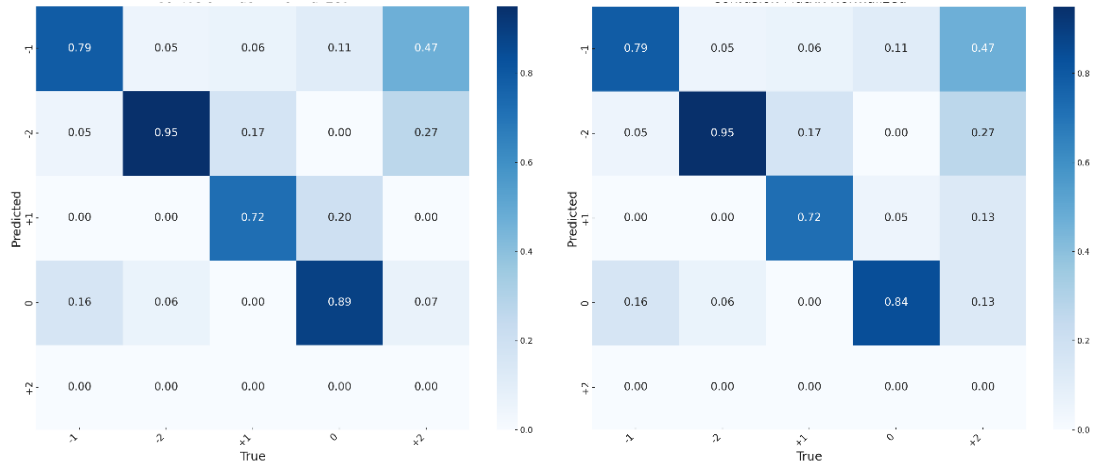


Figure B-1 The normalized confusion matrix in training and validation for Subject 1.

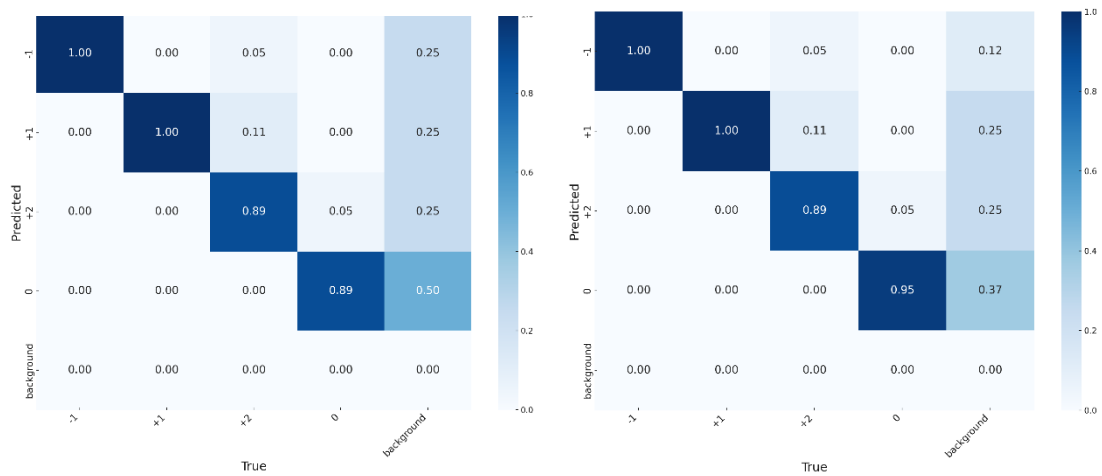


Figure B-2 The normalized confusion matrix in training and validation for Subject 2.

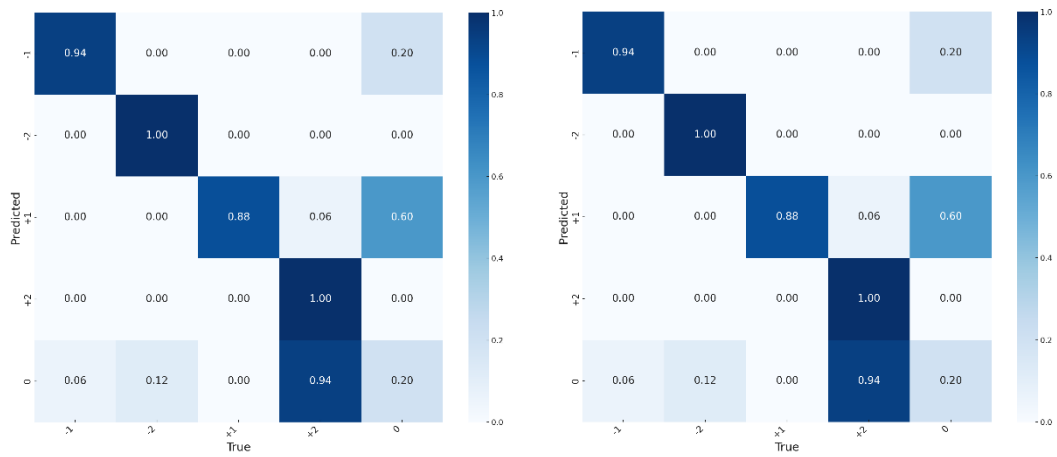


Figure B-3 The normalized confusion matrix in training and validation for Subject 3.

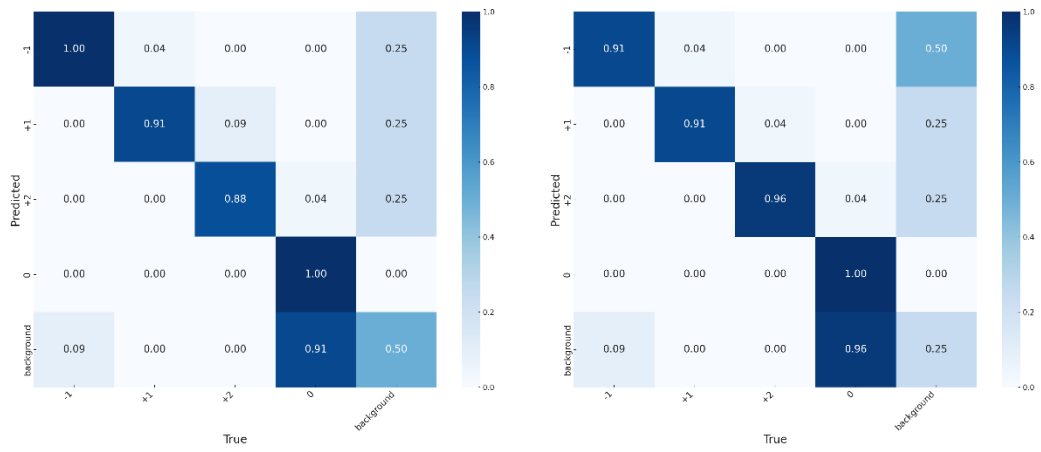


Figure B-4 The normalized confusion matrix in training and validation for Subject 4.

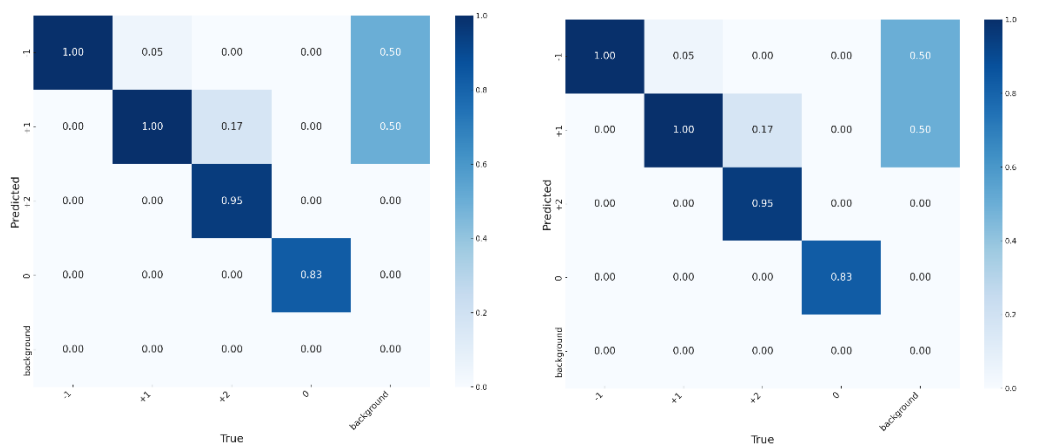


Figure B-5 The normalized confusion matrix in training and validation for Subject 5.

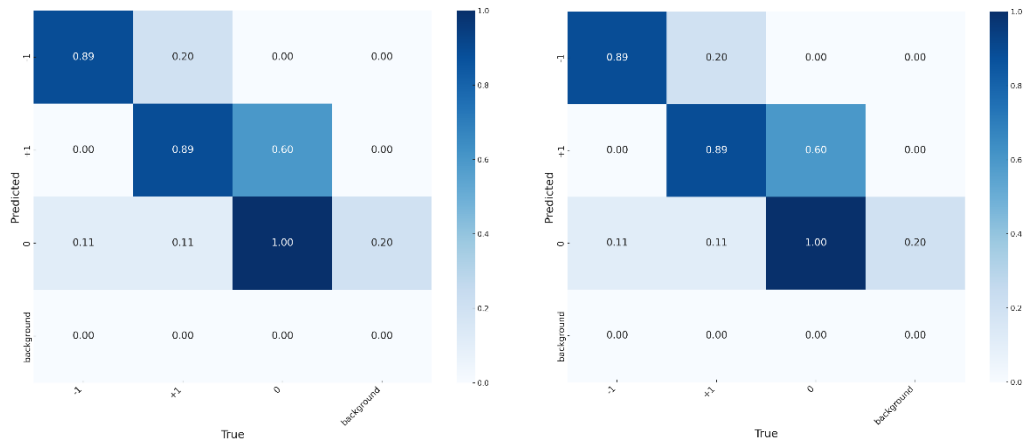


Figure B-6 The normalized confusion matrix in training and validation for Subject 6.

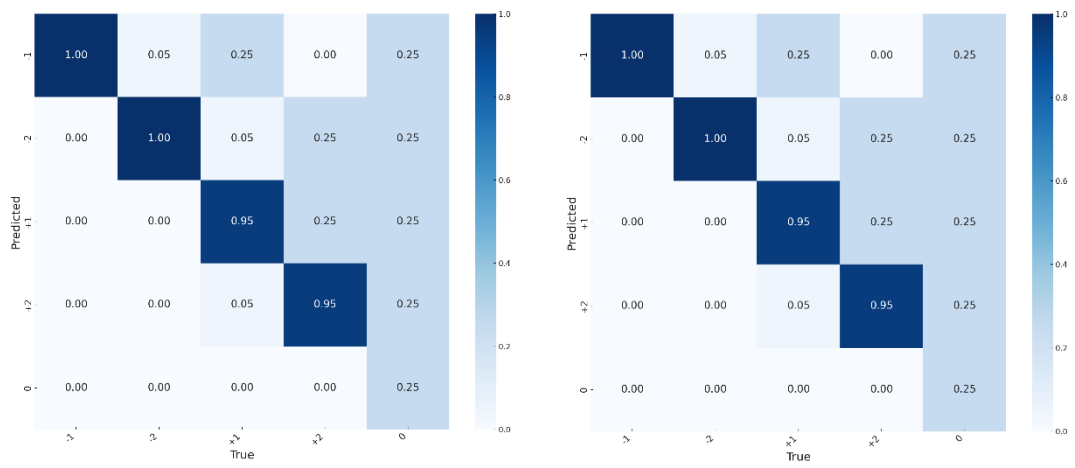


Figure B-7 The normalized confusion matrix in training and validation for Subject 7.

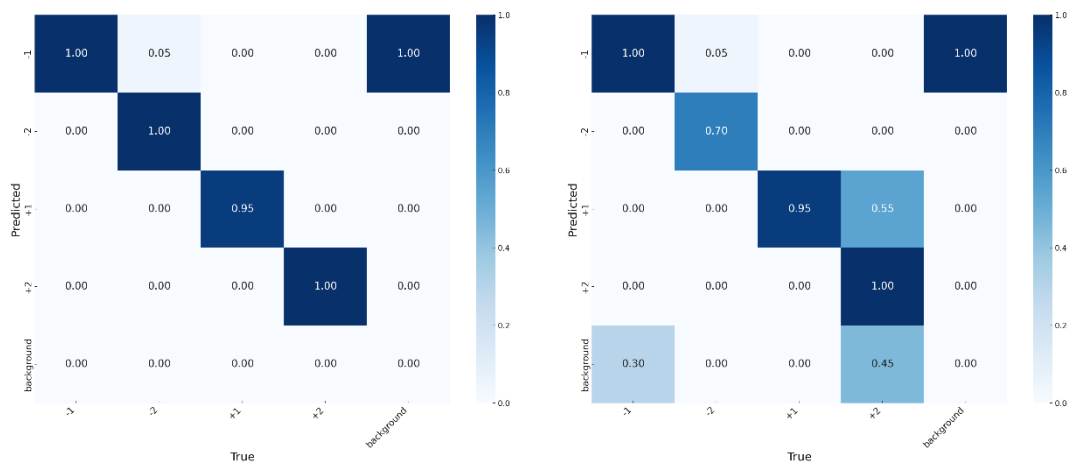


Figure B-8 The normalized confusion matrix in training and validation for Subject 8.

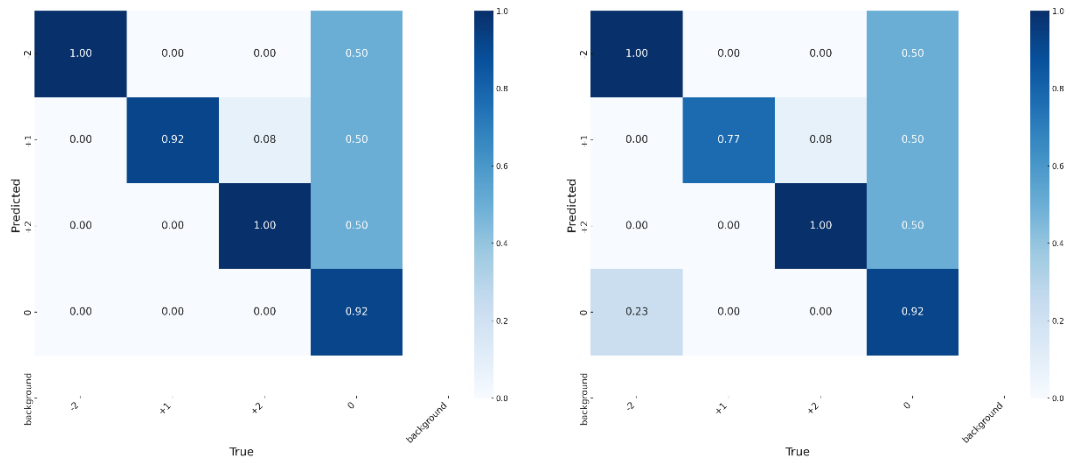


Figure B-9 The normalized confusion matrix in training and validation for Subject 9.

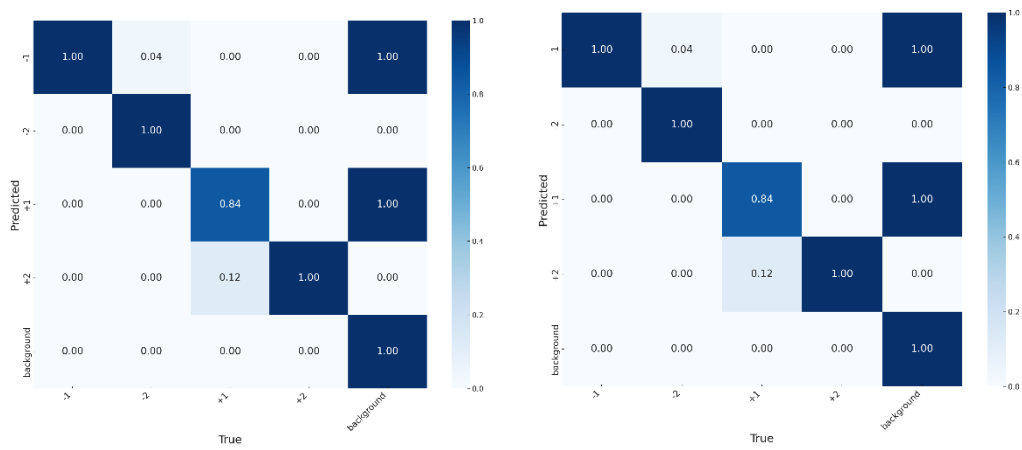


Figure B-10 The normalized confusion matrix in training and validation for Subject 10.

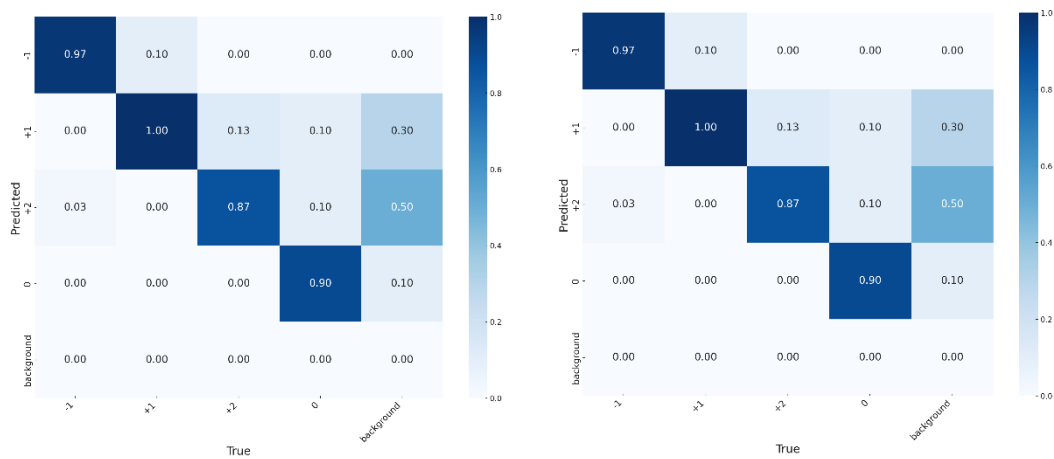


Figure B-11 The normalized confusion matrix in training and validation for Subject 11.

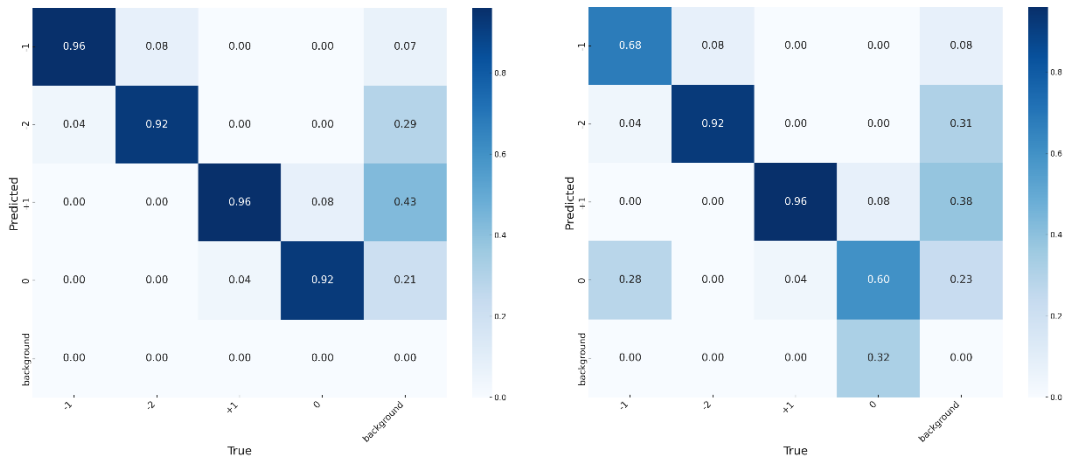


Figure B-12 The normalized confusion matrix in training and validation for Subject 12.

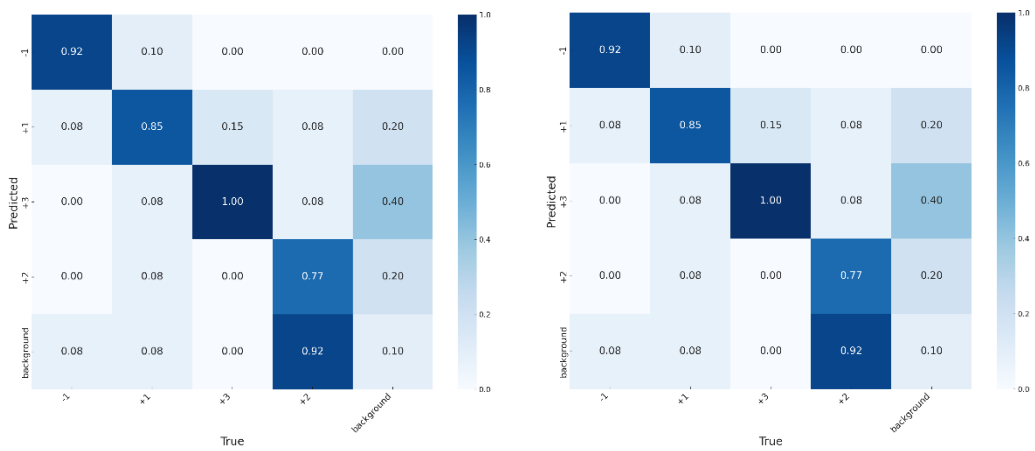


Figure B-13 The normalized confusion matrix in training and validation for Subject 13.

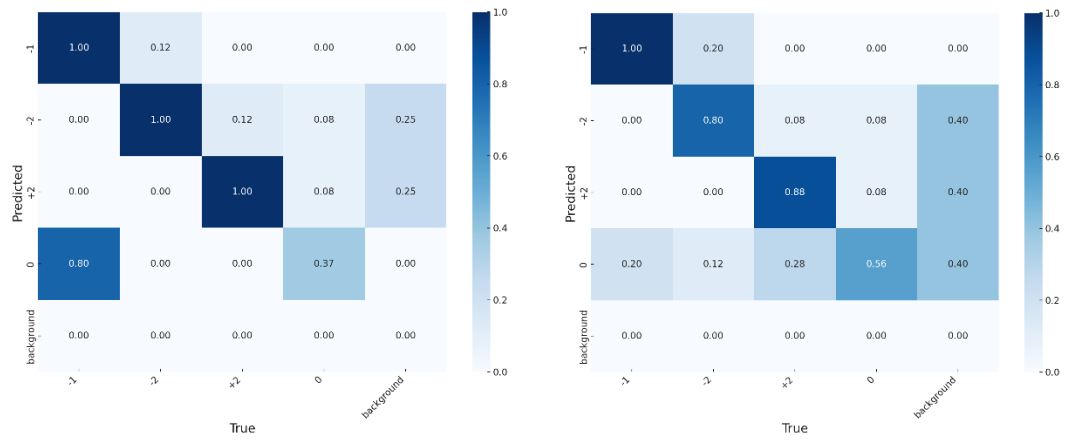


Figure B-14 The normalized confusion matrix in training and validation for Subject 14.

Appendix.C .

This appendix presents the normalized confusion matrix in training and validation for 6 cross-subject experiments in Chapter 5.

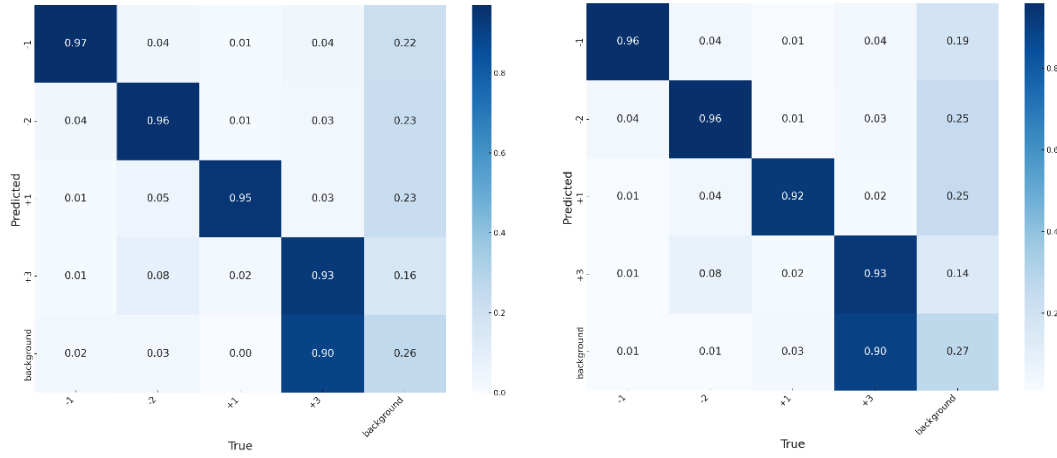


Figure C-1 The normalized confusion matrix in training and validation with the cross-subject dataset from subjects 1-4 and 7-14.

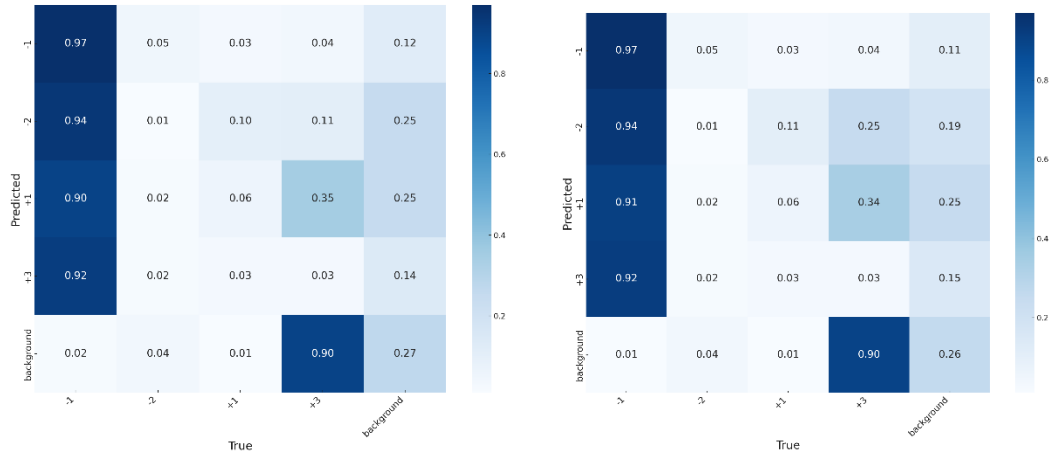


Figure C-2 The normalized confusion matrix in training and validation with the cross-subject dataset from subjects 1-8 and 11-14.

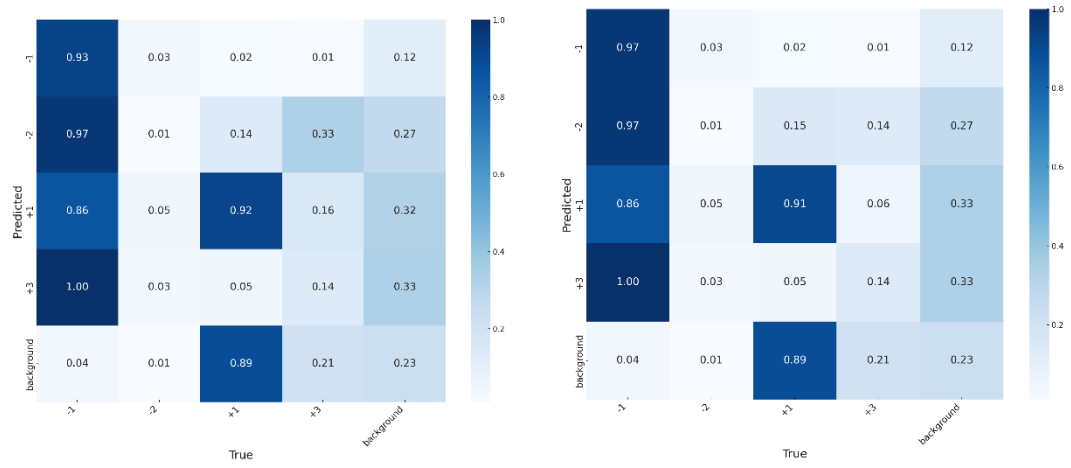


Figure C-3 The normalized confusion matrix in training and validation with the cross-subject dataset from subjects 1-10 and 13-14.

Appendix.D .

This appendix presents the questionnaire for the participants in the experiment in Chapter 5.

Dear participant,

Thanks for agreeing to the test of the thermal image-based personal comfort model. Before the test, we would like to introduce you to the PMV index as standard thermal comfort surveys ask subjects about their thermal sensation on a seven-point scale from cold (-3) to hot (+3) as shown in Figure 1.

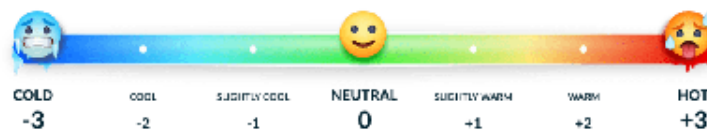


Figure 1 The seven points PMV index from -3 to +3

The test is in a climate chamber and the initial comfort level will be asked when entering the room. Your thermal image will be recorded by a thermal camera during the test and no personal information will be taken. A hotwire and environment sensor will also be used to record the temperature, relative humidity, CO2 level and airspeed.

The heater will be on to raise the temperature till the highest temperature heater can reach. Then the heater will be off and cooling on to explore the thermal comfort in a larger range. Please indicate your thermal comfort level every 5 mins as shown in the form below.

No personal image or information will be taken and please do indicate if there are any uncomfortable in the test and we can stop any minute if you want.

Appendix.E .

Declaration

I declare that the thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly authored publications has been included. My contributions and those of the other authors to this work have been explicitly indicated below.

I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others. The work presented in some of the content described in Chapters 2-5 was previously published in journals [*Renewable and Sustainable Energy Reviews*] and [*Journal of Building Engineering*]s by authors Wuxia Zhang, John Kaiser Calautit, Paige Wenbin Tien, Shuangyu Wei, and Yupeng Wu, and this study was conceived by all of the authors.

Declaration in Chapter 2

Some work presented in this Chapter was previously published in the journal [*Renewable and Sustainable Energy Reviews*] as titled *A Review on Occupancy Prediction Through Machine Learning for Enhancing Energy Efficiency, Air Quality and Thermal Comfort in the Built Environment* by author Wuxia Zhang and co-authors Yupeng Wu and John Kaiser Calautit. I played a major role in Conceptualization, Methodology, and Writing - the original draft and this study were conceived by all the authors.

Declaration in Chapter 4

Some work presented in this Chapter was previously published in the journal [*Journal of Building Engineering*] as titled *Deep Learning Models for Vision-based Occupancy Detection in High Occupancy Buildings* by author Wuxia Zhang and co-authors John

Kaiser Calautit, Paige Wenbin Tien, Yupeng Wu and Shuangyu Wei. I played a major role in Conceptualization, Methodology, and Writing - the original draft and this study were conceived by all the authors.