Modelling Interior Component Stocks of UK Housing Using Exterior Features and Machine Learning Techniques

Menglin Dai*[1], Jakub Jurszyk†[1,2], Charles Gillott[2], Kun Sun[3], Maud Lanau[4], Gang Liu‡[1,5], and Danielle Densley Tingley[2]

[1]College of Urban and Environmental Sciences, Peking University
[2]School of Mechanical, Aerospace and Civil Engineering, The University of Sheffield
[3]Department of Green Technology, University of Southern Denmark
[4]Department of Architectural and Civil Engineering, Chalmers University of Technology
[5]Institute of Carbon Neutrality, Peking University

Abstract

Building stock modelling is a vital tool for assessing material inventories in buildings, playing a critical role in promoting a circular economy, facilitating waste management, and supporting socio-economic analyses. However, a major challenge in building stock modelling lies in achieving accurate component-level assessments, as current approaches primarily rely on archetype-based statistical data, which often lack precision. Addressing this challenge requires scalable methods for estimating the dimensions of interior components across large building stocks. In this study, we introduce the UKResi dataset, a novel dataset containing 2,000 residential houses in the UK, designed to predict interior wall systems and room-level spatial configurations using exterior building features. Benchmark experiments demonstrate that the proposed approach achieves high predictive performance, with an R^2 score of 0.829 for interior wall length and up to 0.880 for bedroom counts, 0.792 for lounge counts, and 0.943 for kitchen counts. Contributions of this work also include the introduction of a multi-modal approach into the field of building stock modelling, integrating exterior features and facade imagery. Furthermore, we analyse the driving factors influencing wall length and room predictions using permutation importance and SHAP values, providing insights into feature contributions, especially facade opening information being a critical driving factor of modelling interior features. The UKResi dataset serves as a foundation for future component-level building stock modelling, offering a scalable and data-driven solution to assess building interiors. This advancement holds significant potential for improving material inventory assessments, enabling more accurate resource recovery, and supporting sustainable urban planning.

Keywords: building stock modelling, urban sustainability, circular economy, machine learning, households, building material

*Corresponding author, Email: menglin.dai@pku.edu.cn
†This author contributed equally to this work.
‡Corresponding author, Email: gangliu@pku.edu.cn

1 Introduction

In the global pursuit of meeting the 1.5°C target set by the Paris Agreement and advancing the United Nations Sustainable Development Goals (SDGs) (United Nations, 2015; United Nations Framework Convention on Climate Change (UNFCCC), 2015), particularly Goal 11 (Sustainable Cities and Communities) and Goal 12 (Responsible Consumption and Production), residential buildings play a pivotal role in reducing embodied and operational carbon emissions. Low-rise residential buildings ($\leq$ 3 storeys) are particularly significant in this context. In the United Kingdom, these structures constitute 79% of the total housing stock count (Ministry of Housing, Communities & Local Government, 2018) and accommodate 78% of households (Office for National Statistics, 2023). Meanwhile, they are responsible for extensive material (e.g. clay, sand and timber) and energy consumption, resulting in significant embodied and operational carbon emissions throughout their lifecycle (Z. Cao et al., 2020; Heeren et al., 2015; Zhong et al., 2021, 2022). Low-rise residential buildings have also been identified as the primary contributor to housing material stock accumulation in countries such as Austria (Haberl et al., 2021) and the United States (Frantz et al., 2023). Therefore, understanding the mass composition, energy consumption and spatial distribution of these buildings is critical for climate change mitigation and the advancement of a circular economy. This is particularly pertinent in the UK, in light of the targeted delivery of over 300,000 homes per year (UK Government, 2024).

Among the state-of-the-art building material stock accounting methods, the bottom-up approach serves as a fundamental tool for estimating the material mass stocks at the building level (Lanau et al., 2019; Pei et al., 2024). This method assumes building material composition to be homogeneous within a predefined archetype, multiplying associated material intensity coefficients by building geometries (e.g. floor area or volume) to calculate mass stock.

Accuracy in the extraction of dimensional exterior building features and the calculation of material intensity coefficients are key factors in the bottom-up approach. Owing to advancements in remote sensing and machine learning technologies, methodologies for extracting exterior building features (e.g. height (Cai et al., 2023; Y. Cao & Weng, 2024) and footprint (Buyukdemircioglu et al., 2022; Guo et al., 2022)) at scale are increasingly being explored. Simultaneously, the widespread availability of point-cloud data has made precise modelling of building envelopes feasible (Q. Hu et al., 2021; Krapf et al., 2023), with pre-processed inventory datasets for use in building stock modelling also growing in availability (Milojevic-Dupont et al., 2023).

Despite increasing insight on national material intensities and generalised international databases (Fishman et al., 2024; Lanau & Liu, 2020), the assumption of homogeneous material distribution within the same predefined archetype has been questioned in a number of studies. These have found significant discrepancies in material intensity within the same archetype (Arceo et al., 2021, 2023; Miatto et al., 2023; Nasiri et al., 2023), potentially leading to substantial errors in subsequent material accounting. Currently, material intensities are also typically reported at the aggregated-material level (i.e. ‘steel’) rather than component form (i.e. ‘steel beams’), failing to provide sufficient compositional information for required insights on associated reuse and recycling potential. Whole-building circular economy potential is similarly neglected, with a typically limited consideration of the configuration (e.g. within a wall, floor or roof) and provided function (e.g. as part of a bedroom, kitchen or bathroom) of existing residential material stocks.

Such an effect is worsened by stock studies' focus on the 'structure' and 'skin' layers, often overlooking the inventory of (semi-)permanent interior items (e.g. radiators and plumbing, kitchens and sanitary-ware and appliances and furniture) in the 'services','space' and 'stuff' layers (Brand, 1995). In addition to their variable composition, the primary challenge in accounting for these items using the bottom-up approach lies in their heterogeneity within archetypes and individual buildings and resultantly poor scaling with floor area or volume. For this reason, recent research on interior residential stocks predominantly relies on top-down statistical data (Arora et al., 2019; X. Li et al., 2023; Liu et al., 2020) and/or focuses on fast moving consumer goods (e.g. food and clothing) (Di Donato et al., 2015; Kissinger & Damari, 2021).

Where residential structural material stocks may be estimated using external features and geometries, estimating the quantity, form, configuration and function of (non-)structural residential materials requires nuanced understanding of interior features. Understanding building interiors at large scale is challenging due to the difficulty of data acquisition. Physical surveys of every building in a city or nation to ascertain interior layouts would be prohibitively time consuming and thus cost intensive. Thus research has explored ways of predicting internal features from external images. This includes attempts to understand the interior space of buildings include techniques such as interior image segmentation (Zhou et al., 2019), interior scene reconstruction (Budroni & Boehm, 2010) and consequent BIM (Building Information Modelling) model auto-generation (Mahmoud et al., 2024), though the necessary data, i.e. from the interior, remains largely inaccessible. Huang et al. use facade images to predict the floor area of houses, suggesting a promising direction for employing more readily available exterior features to infer interior details (Huang et al., 2024). In our previous research, considering a preliminary dataset of 300 samples, we investigated the potential of using exterior house features to predict the length of interior walls in UK housing (Dai et al., 2024). Although estimating the quantity, form and configuration of both structural and non-structural wall materials, this did not provide insight on their function nor facilitate consideration of (semi-)permanent interior items in the 'services','space' and 'stuff' layers (Brand, 1995).

This study builds upon our previous work by developing the multi-modal 'UKResi' dataset, containing internal and external imagery (e.g. facade and room interior), geometry (e.g. building width/depth and wall length) and labelled features (e.g. room function and window/door counts) for 2,000 houses in the UK. Utilising the UKResi dataset, we apply a range of multi- and single- modal machine learning techniques to successfully predict a number of internal building features (e.g. wall length and room counts) from external images and derived attributes (e.g. building width and depth). Following this, we categorise exterior features according to their availability and prediction importance, providing further insight on the potential for large-scale residential interior stock prediction under a range of different scenarios.

2 Materials and Methods

2.1 Dataset Construction

Data Collection and Annotation Process The constructed UKResi dataset consists of 2,000 housing units from England and Wales. These samples were acquired through Zoopla, a real estate platform established in 2007 that ranks among the largest property search websites in the UK (Hancock, 2022). The dastaset documents housing units with facade images, interior photographs, floor plans, and geographical locations. Three authors with architectural expertise formed the annotation team. To maintain consistency, only one member performed the dataset annotation, while the other two independently verified consistency. This approach ensured consistent labelling without the need for an inter-annotator agreement.

A significant obstacle in utilising this data was the variability in floorplan quality, as the information is sourced from various agencies, each providing data with differing levels of quality. To address this, the authors, who possess architectural expertise, were tasked with verifying the quality of the samples before proceeding with the annotation process. The procedure for data acquisition also took into account the geographical location of properties. Attention was given to guarantee an approximately even distribution of the 2,000 samples across the nine regions of England and Wales.

Figure 1 illustrates the comprehensive annotation methodology employed in developing the UKResi dataset. The dataset encompasses four categories of data extracted from the Zoopla website: location details, facade images, interior photographs, and floorplans. Initially, data collection involved the local download of relevant information which was then organised into separate folders corresponding to each category of data - excluding the description. The latter, displayed on the website, was methodically recorded in a CSV (Comma-Separated Values) file.

The annotation process was performed using Computer-Aided Design (CAD) software and comprised three stages: 1) scaling, 2) measuring, and 3) quality assessment. Scaling consisted of importing the floorplan data into the CAD system to use the scaling tool and the labelled dimension of the floorplan to bring the floorplan to its actual dimensions. In the measuring phase, CAD's ruler tool was used to measure all pre-specified features. Quality assessment was conducted during data collection, with authors possessing architectural expertise visually inspecting the floorplan, and at the conclusion of the annotation process. The data screening process involved using facade and interior images as ground truth of the building to compare against the floorplan layouts. For example, if the facade image shows two windows on the ground floor but the floorplan indicates only one, that sample was rejected. Similarly, if an interior image shows an open-plan kitchen and dining room, but the floorplan depicts these spaces as separate, the sample was excluded. This rigorous approach ensured that only accurately matched floorplans were retained. In this article, the term 'ground floor' refers to the building's lowest level at ground level, and 'first floor' denotes the level immediately above the ground floor. In the final quality assessment phase, external feature dimensions were used to cross-verify the accuracy of measurements. Facade imagery was applied to estimate window and building dimensions using elements of known sizes, such as bricks, and these estimates were compared with those obtained from floorplans. The early and the late quality assessments ensured the reliability

Figure 1: The schematic of the dataset construction process: Part-A involves the identification of appropriate samples by examining their data completeness, types and geographical locations; Part-B presents sample examples which encompass images of facades and interiors, floorplans, and descriptive data; Part-C depicts the data annotation procedure, which incorporates the use of scaling and ruler tools; Part-D indicates the compilation of this data into a CSV file.

of both data and measurement quality. Worth noting is the use of Google Street View for quality assurance in cases where the facade image from the Zoopla website was not usable due to too large viewing angles or obstructions.

Feature Definitions and Annotation Details In total, 38 features were identified and documented or annotated. These features encompass extensive information retrievable from the gathered multi-modal data. Depending on the source of the label extraction, the thirty-eight features were categorised into ten exterior features related to dimensions, twenty-one interior features, and seven property features related to building attributes (e.g., building energy performance labels). These attributes are summarised with their definitions in the supporting information, SI-attributes_summary.

Multiple features were used to properly describe the shape irregularities of building footprints. Beyond typical features such as building width, depth, area, and perimeter, the short depth was annotated (i.e., the length of the house's shorter side). The average depth was also registered, which reflects the area-standardised building depth and is calculated by dividing the gross area of the building by its facade width, as shown in Figure 2-I. Additionally, Figure 2-II documents the width and count of the building facade openings, encompassing both windows and doors.

The set of 21 interior features was designed to thoroughly account for wall attributes that offer structural support and divide the interior into distinct zones and their respective functions, i.e. bedroom, kitchen, etc. Within these features, an interior wall was categorised as either a main inner wall or a storage wall. As illustrated in Figure 2-II, the main inner wall includes both load-bearing and partition walls, as the floorplan data did not facilitate distinctions between the two. In contrast, the storage wall, which creates a permanent storage area, could be identified through inspections of both floorplans and interior images.

Doors, a critical component in the interior space that provides control of space combinations, was characterised through 11 features. As shown in Figure 2-III, four different types of opening were defined, including standard door, non-standard door, storage door, and non-door opening. For each door type, their widths were also measured, in addition to door counts.

The dataset specifies eight distinct zoning function attributes, encompassing bedroom, kitchen, bathroom, toilet, lounge, small room, dining room, and total number of spaces. Small rooms were identified as compact individual spaces often used for wardrobes, or offices. The distinction between bathrooms and toilets was based on the presence of shower facilities. Furthermore, three connection features—living-dining, kitchen-dining, and kitchen-living—were included to characterise open-plan designs found in some buildings, illustrating varied space functionalities. Examples can be seen in Figure 2-III. The total room count represents the aggregate of each separate room. The seven property features include number of floors, building attachment type, price, energy label, location, form type (whether the sample is a house or a bungalow) and whether a loft conversion exists.

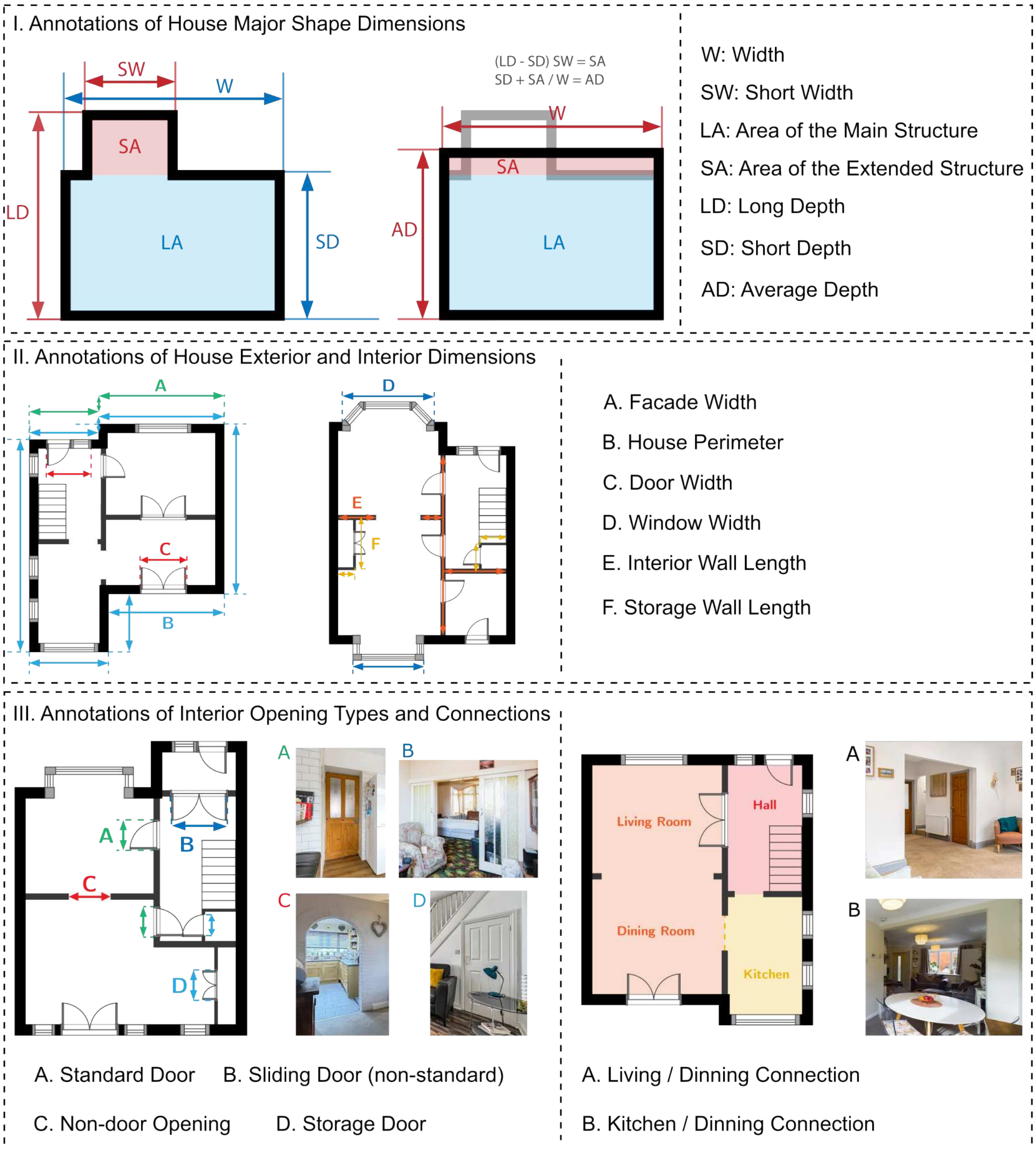

Figure 2: Illustrations of annotations for various features. Panel I shows the methods used to determine width and depth. Panel II illustrates the definition of opening, perimeter, and interior wall length. Panel III illustrates the annotation of interior openings.

2.2 Dataset Summary

The built dataset contains 2,000 individual buildings with 3,859 floors. Their spatial distributions are presented in the supporting information, SI-Data_Distributions. The 2,000 samples are evenly distributed across the nine regions of England and Wales. We also conduct a comparative analysis between the building type compositions and overall floor areas of the constructed dataset and those derived from the English Housing Survey (Department for Communities and Local Government, 2016; Ministry of Housing, Communities & Local Government, 2010) in the same supporting information document. The result demonstrates that when comparing the two data sources—UKResi and the English Housing Survey—across the four house types (detached, semi-detached, terraced, and bungalow) and three gross area categories ($< 70sq.m., 70 - 89sq.m., > 90sq.m.$), while subtle distinctions exist, the two datasets present broadly consistent compositions: For detached houses, both sources report a strong concentration in the $> 90sq.m.$ category, though UKResi shows a slightly higher proportion (2%). Semi-detached houses in UKResi are somewhat more skewed toward larger sizes, with 12% differences between 70-89 sq.m. and $> 90sq.m.$, than those in the UK Housing Survey, which appear evenly distributed. Terraced homes, while generally leaning to the larger category in UKResi, another 12% difference, show a 5% greater share in the mid-range sizes according to the UK Housing Survey. Bungalows exhibit a notable contrast, with UK Housing Survey data placing a 12% more share in the smallest size category than UKResi. Despite these differences in emphasis, the overall patterns are broadly similar: both datasets suggest that Detached and Terraced homes tend to be larger, while Bungalows are more commonly smaller, and Semi-Detached units often fall between these extremes.

2.3 Benchmark Experiment

Benchmark Pipeline and Model Selection To establish a unified framework for advancing the house interior prediction task, benchmark experiments were developed to deploy various machine learning models on the constructed UKResi dataset. These tests ensured that all methods were assessed using identical data and metrics, thereby fostering fairness, reproducibility, and advancement. By consistently evaluating performance, we can confidently rely on improvements, spotlight leading solutions, and inform future investigations. Essentially, benchmarks uphold standards and drive the discipline forward which has been validated as essential in machine learning research (Deng et al., 2009).

This dataset was primarily designed to facilitate scalable estimation of building interior inventories, encompassing both wall components and enduring items. To achieve this objective, nine key features were established as the benchmark targets: namely, the length of interior walls (serving as an indicator of the wall inventory) and the counts of eight distinct spaces (reflecting the stock of enduring items stated in Section 2.1). In the context of scalable estimation, the primary challenge lies in the availability of data. Consequently, when using the constructed dataset on a large scale, the accessibility or practicability of extracting annotated exterior features becomes crucial to consider. Initially, we categorised the annotated exterior features into three distinct levels based on their accessibility, as indicated in Table 1. The classification demonstrated that features derivable from the building footprint were designated as high level. Features obtainable from facade images necessitating an image classification model were labelled as medium level, while those requiring object detection or segmentation for extraction were identified as low level. Subsequently, the access hierarchy could serve as a tool to evaluate the influence of external features on internal prediction. By assessing feature impacts based on different levels of accessibility, the minimum cost of applying the method at a large scale can be evaluated which significantly contributes toward the scalability of the method.

Access Level	Features	Accessibility
High	area, long depth, short depth, average depth, width, perimeter	footprint data e.g. Google(Google Research, 2021), Bing(Microsoft, 2024), OS MasterMap(Ordnance Survey, 2023)
Medium	attachment type, form type	facade image e.g. street view services (Anguelov et al., 2010) and classification model (Dai, 2023)
Low	counts and widths of facade windows and doors, number of floors	facade image and detection model (H. Li et al., 2023)

Table 1: The table presents the defined exterior feature accessibility hierarchy and the corresponding data sources or methodologies for acquiring. High-level features can be derived directly from footprint data. Medium-level features require facade images and a classification model. Low-level features demand more advanced image processing, including detection models applied to facade images.

This benchmark test was constructed using a multi-task learning framework because the given problem required predicting multiple outputs concurrently. Multi-task learning is a machine learning paradigm in which related tasks are learned simultaneously, allowing knowledge gained from one task to benefit others. This approach can significantly reduce computational overhead and improve model generalisation (Zhang & Yang, 2021). In the context of the interior prediction task, forgoing a multi-task setup would necessitate training nine separate models, substantially increasing both computational cost and deployment complexity.

In addition to multi-task learning, this benchmark also employed a multi-modal learning technique. Multi-modal learning, a rapidly advancing area of machine learning research, integrates multiple data types—such as images and text—into a single model. By merging complementary information sources, multi-modal models can learn richer, more nuanced representations than single modality can provide (Ramachandram & Taylor, 2017). This approach has been shown to enhance prediction performance in numerous applications (Xu et al., 2023). For the interior prediction task specifically, we integrated facade imagery with annotated structured data. Since facade images typically contain rich building-related information, including age and construction style, we hypothesised that combining these visual cues with structured data will lead to improved interior prediction performance.

Figure 3 illustrates the benchmark experiment setup and the corresponding models. Part A presents the feature combinations according to the accessibility hierarchy defined in Table 1. For the single-modal tests, three common machine learning models—Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and Random Forest (RF)—were adopted (Part C). Each model was trained on each set of feature to systematically evaluate how removing varying accessibility levels of exterior attributes affected the performance of interior feature estimation. Part D depicts the proposed multi-modal model, FacIntNet. FacIntNet includes a feature extraction network that processes facade images, applying a 1×1 convolution to reduce the number of channels from the extracted features.

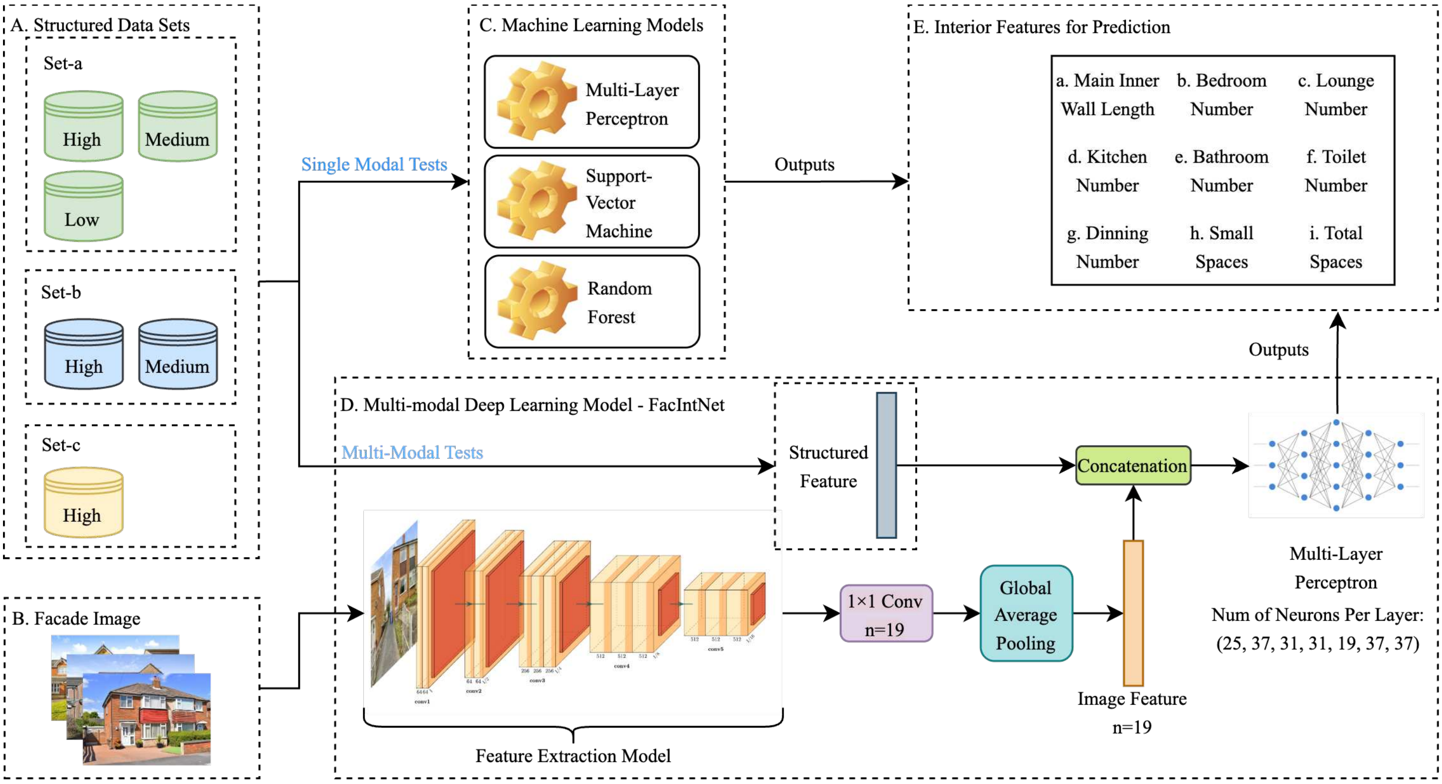

Figure 3: The figure illustrates the pipeline for benchmark tests and the architecture of the developed multi-modal deep learning model. According to the feature access tiers outlined in Table 1, exterior features are organised by progressively removing lower access level features, as depicted in Part A. Subsequently, these three generated datasets are benchmarked using three widely adopted machine learning models, as shown in Part C. These distinct datasets are then integrated with facade image data to assess the potential influence of facade images on predicting interior features in Part D.

A global average pooling layer (Lin, 2013) then compresses the resulting feature maps into a vector. This vector is concatenated with the structured data vector, and the combined features are input into an MLP to generate the final predictions. The design of FacIntNet aims to determine whether incorporating facade image features can improve the accuracy of interior predictions. Three different feature extraction networks were adopted including ResNet50 (He et al., 2016), ResNetV2 (Szegedy et al., 2017) and Xception (Chollet, 2017). All three models have demonstrated strong feature extraction capabilities and have achieved significant success (Chen et al., 2018; J. Hu et al., 2018; Woo et al., 2018).

Training Configurations and Results Evaluation All benchmark tests were performed using Python on a workstation with a Linux Ubuntu 22.04 operating system, an Intel Xeon Silver 4310 processor, a Nvidia RTX 4090 graphics card and 64 gigabytes(GB) of RAM. Initially, the 2,000 samples were randomly divided into training and validation sets based on their building IDs in an 80%:20% ratio, a standard choice in machine learning, utilising a random seed of 28. Subsequently, the floor data associated with each building ID was clustered into their respective sets to construct the training and validation datasets.

The single-modal models were implemented using the scikit-learn library(Pedregosa et al., 2011). A grid search procedure was employed to determine optimal hyper-parameters for each model. Specifically, for the Multi-Layer Perceptron (MLP), the number of layers ranged from 3 to 8 with 10 to 50 neurons per layer. For the Random Forest (RF), the number of estimators ranged from 10 to 40 and the maximum tree depth from 10 to 50. Due to the limitation of the Support Vector Machine (SVM) model, individual models for each attribute were built for the SVM test, the search explored various regularisation coefficients, C including 0.1, 1, 10, 100 and 1000, kernel functions, rbf or poly, epsilon values including 0.01, 0.1, 0.5 and 1.0, and gamma values, scale or auto.

The multi-modal models were implemented using TensorFlow (Abadi, 2016). Training configurations included a batch size of 16 and an initial learning rate of 0.0001, which was halved every 10 epochs, if without loss decrease, until reaching 0.00001. The models used the Adam (Kingma & Ba, 2014) optimiser. Input images were padded to square shape and resized to 512×512 pixels, and data augmentation—consisting of a 0.1 spatial shift and 50% chance of horizontal flips—was applied. Each model was trained for 500 epochs with an early stopping setting of 50 epochs.

All models were evaluated using three common regression metrics: R^2, RMSE, and MAE. The R^2 score measures the correlation between predictions and ground truth, while RMSE quantifies the average magnitude of errors, placing greater emphasis on larger deviations. MAE measures the average magnitude of errors on a linear scale. For outputs related to the number of spaces and different rooms, predictions were rounded to the nearest integer before computing these metrics. The functions used for calculating these three metrics are listed below, where y_i is the observed value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the observed values and n is the number of observations.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{2}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3}$$

Furthermore, Permutation Importance (Altmann et al., 2010) and SHapley Additive exPlanations (SHAP) (Lundberg, 2017) values were used to quantify the contributions of each individual exterior feature to the predictions of interior features. Permutation Importance evaluates the significance of a feature by measuring the impact of randomly shuffling its values on the model's performance. This method helps determine the extent to which the model relies on a particular feature for making accurate predictions. In contrast, SHAP values provide a detailed explanation of a model's output by assigning each feature a numerical value representing its contribution to a specific prediction (sample-based). SHAP offers insights into both the magnitude and direction of a feature's influence on the prediction. Finally, to enhance the interpretability of the model's behaviour, the top five best-performing and least-performing examples were visualised. This visualization facilitates a deeper understanding of how the model makes predictions and responds to different input features.

3 Results

3.1 Benchmark Test Results

Table 2 demonstrates the results of the feature-based ablation study for three machine learning models: Multi-Layer Perceptron (MLP), Random Forest (RF), and Support Vector Machine (SVM). The highest R^2 score for each feature set exceeding the defined 0.6 threshold is highlighted in bold, serving as an indicator of a good fit. As shown in the table, the model performance generally declines as the number of exterior features decreases, moving from the High + Medium + Low feature set to the High + Medium set, and finally to the High set. This trend is particularly evident in the prediction of interior wall length and total room count, where the R^2 score of the interior wall length prediction drops from 0.842 when using all defined exterior features, to 0.770 after removing low-level features, and further to 0.749 when only high-level features are retained. For total room count predictions, the R^2 score decreases from 0.661 when using all features, to 0.490 when facade features are removed, and finally to 0.460 when only footprint features are used. In the predictions of other room counts, the performance decline hierarchy is not as visible as the two aforementioned attributes. This is mostly due to starting from removing low-level features, room counts are not able to properly predicted leading to similar outcomes in High + Medium and High sets as shown in Table 2. In addition, for the toilet count and small room count, none of the trained models achieve an R^2 score above the 0.6 threshold across all feature sets. This suggests that the variability or randomness in the configurations of these room types may be contributing to the models' poor performance.

Among the three selected models, the Multi-Layer Perceptron (MLP) consistently achieves the best performance for predicting interior wall length and total room count across all feature sets. Specifically, the MLP achieves an R^2 score of 0.842 for wall length prediction and 0.661 for total room count prediction when all features are included. The Random Forest (RF) model follows closely, achieving the second-best performance for these metrics. In contrast, the Support Vector Machine (SVM) shows the poorest performance for predicting wall length and total room count. However, when predicting the number of specific rooms, including bedrooms, lounges, kitchens, and dining rooms, the SVM model outperforms the MLP and RF models. For instance, the SVM achieves the highest R^2 scores of 0.878 for bedroom count, 0.792 for lounge count, and 0.933 for kitchen count when using all features. This indicates that the SVM is more effective at handling tasks involving distinct, discrete categories of rooms.

Examining the error metrics, such as the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), further confirms the decline in model performance as the number of exterior features is reduced. For example, the RMSE for the MLP's wall length prediction increases from 2.885 with all features to 3.642 when only high-level features are used. This trend underscores the importance of exterior facade features in maintaining model accuracy. Removing these features significantly impacts the models' ability to predict interior attributes, particularly for room counts.

These results highlight the strengths and limitations of each model. The MLP is robust for predicting aggregate

Feature Sets	Model	Metric	Interior Attributes								
			Wall	Bed	Lounge	Kitchen	Dinning	Bath	Toilet	Small	Total
High + Medium + Low	MLP	R^2	**0.842**	0.854	0.757	0.923	0.670	0.661	0.046	0.028	**0.661**
		RMSE	2.885	0.552	0.283	0.140	0.316	0.415	0.450	0.442	0.862
		MAE	2.167	0.239	0.075	0.020	0.097	0.169	0.202	0.185	0.564
	RF	R^2	0.833	0.876	0.784	0.933	0.688	0.677	0.312	-0.004	0.657
		RMSE	2.967	0.507	0.266	0.131	0.307	0.405	0.382	0.450	0.866
		MAE	2.223	0.213	0.068	0.017	0.092	0.164	0.146	0.186	0.549
	SVM	R^2	0.805	**0.878**	**0.792**	**0.933**	**0.692**	**0.684**	0.325	0.087	0.653
		RMSE	3.212	0.503	0.261	0.131	0.305	0.400	0.378	0.429	0.872
		MAE	2.170	0.201	0.066	0.017	0.091	0.160	0.143	0.176	0.550
High + Medium	MLP	R^2	**0.770**	0.056	-0.289	-0.417	-0.367	-0.183	-0.394	-0.037	0.490
		RMSE	3.487	1.402	0.651	0.601	0.643	0.774	0.543	0.457	1.056
		MAE	2.712	1.247	0.416	0.356	0.403	0.584	0.295	0.201	0.761
	RF	R^2	0.764	0.053	-0.233	-0.397	-0.302	-0.154	-0.401	-0.083	0.476
		RMSE	3.534	1.404	0.637	0.596	0.627	0.765	0.545	0.467	1.071
		MAE	2.806	1.239	0.398	0.350	0.386	0.561	0.297	0.205	0.774
	SVM	R^2	0.723	0.032	-0.345	-0.618	-0.722	-0.322	-0.425	-0.057	0.460
		RMSE	3.826	1.419	0.665	0.642	0.722	0.819	0.549	0.461	1.087
		MAE	2.789	1.264	0.429	0.407	0.503	0.597	0.302	0.202	0.776
High	MLP	R^2	**0.749**	0.034	-0.285	-0.484	-0.336	-0.170	-0.413	0.009	0.460
		RMSE	3.642	1.418	0.650	0.615	0.636	0.770	0.547	0.447	1.087
		MAE	2.809	1.262	0.417	0.373	0.396	0.575	0.299	0.197	0.777
	RF	R^2	0.749	0.051	-0.269	-0.459	-0.284	-0.172	-0.481	-0.135	0.445
		RMSE	3.641	1.405	0.646	0.609	0.623	0.771	0.560	0.478	1.102
		MAE	2.846	1.243	0.409	0.366	0.383	0.563	0.314	0.213	0.789
	SVM	R^2	0.709	0.007	-0.517	-0.721	-0.527	-0.423	-0.450	0.002	0.415
		RMSE	3.921	1.438	0.706	0.662	0.680	0.850	0.554	0.448	1.131
		MAE	2.848	1.282	0.475	0.433	0.451	0.614	0.307	0.193	0.789

Table 2: The table displays the results of the single-modal benchmark assessment using three prevalent machine learning models: where MLP denotes the multi-layer perceptron, RF symbolises the random forest, and SVM signifies the support vector machine. The highest R^2 score for each feature set test exceeding 0.6 is highlighted in bold.

measures like wall length and total room count, while the SVM excels at predicting individual room counts. The RF model strikes a balance, performing consistently across most metrics. The findings suggest that an ensemble learning model which combines multiple models may further enhance predictive performance. Overall, the study demonstrates that the selection of features plays a crucial role in model performance. Facade features appear particularly important for achieving accurate predictions.

Table 3 presents the results of the ablation study incorporating an additional facade image modality. Three distinct feature extraction networks—ResNet50, ResNetV2, and Xception—are employed within the same architecture to predict interior attributes. A consistent trend of performance drop is observed as the number of input features is progressively reduced from High + Medium + Low to only High-level features, similar to the single-modality results in Table 2. For example, the wall length prediction R^2 score decreases from 0.829 in the High + Medium + Low set to 0.750 in the High + Facade set, highlighting the importance of additional exterior features (e.g., facade details and building types). This trend confirms that the models perform optimally when a comprehensive set of exterior features is included. The predictions for toilet count and small room count continue to fall below the R^2 threshold of 0.6 across all models and feature sets. This further suggests that the randomness or variability in the configurations of these room types makes them inherently difficult to predict.

By cross-comparing both single- and dual- modal experiments, the inclusion of facade imagery does not yield significant improvements in predictions. Slight improvements are observed in the bedroom, kitchen and total spaces count predictions, where these performances improve with the inclusion of the facade image modality. For instance, in Table 2, the SVM model achieves a R^2 score of 0.878 under the High + Medium + Low feature set. In Table 3, the ResNet50 model achieves a slightly higher R^2 score of 0.880 for the same attribute. In similar veins, kitchen prediction increase from 0.933 to 0.943 and total spaces prediction increase from 0.661 to 0.673. This may suggest that the facade image modality provides additional visual cues that are particularly beneficial for these specific room counts predictions. However, for other attributes like interior wall length, bathroom count, the performance generally declines.

When comparing model performance, the ResNet50 model achieves the best results overall in the multi-modal setup. For instance, ResNet50 attains the highest R^2 scores of 0.829 for wall length, 0.880 for bedroom count, and 0.673 for total room count predictions under the High + Medium + Low feature set. The ResNetV2 and Xception models perform slightly worse, with lower R^2 scores and generally higher error metrics. The error metrics, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), further confirm the trends discussed. For example, the RMSE for wall length prediction increases from 3.003 in the High + Medium + Low set to 3.666 in the High + Facade set using ResNet50. This increase in error reflects the loss of predictive accuracy caused by the reduction in features, despite the additional facade modality.

A critical observation arises when examining the facade-only results (bottom rows of Table 3), where no attribute achieves a meaningful R^2 score. For example, the wall length prediction produces R^2 scores as low as -0.021 (ResNet50) and -0.012 (Xception), while predictions for other interior attributes, including room counts and total

Feature Sets	Stem Model	Metric	Interior Attributes								
			Wall	Bed	Lounge	Kitchen	Dinning	Bath	Toilet	Small	Total
High + Medium + Low + Facade	ResNet50	R^2	**0.829**	**0.880**	**0.792**	0.928	**0.692**	0.645	0.281	0.093	**0.673**
		RMSE	3.003	0.501	0.261	0.136	0.305	0.424	0.390	0.427	0.846
		MAE	2.220	0.211	0.066	0.018	0.091	0.180	0.152	0.175	0.554
	ResNetV2	R^2	0.825	0.871	0.788	**0.943**	0.688	**0.671**	0.219	0.074	0.656
		RMSE	3.038	0.519	0.264	0.120	0.307	0.408	0.407	0.432	0.868
		MAE	2.316	0.219	0.067	0.014	0.092	0.167	0.165	0.178	0.562
	Xception	R^2	0.814	0.864	0.764	0.912	0.675	0.671	0.263	0.080	0.645
		RMSE	3.137	0.532	0.278	0.149	0.314	0.408	0.395	0.430	0.881
		MAE	2.318	0.231	0.075	0.022	0.096	0.167	0.156	0.177	0.580
High + Medium + Facade	ResNet50	R^2	0.728	0.064	-0.237	-0.397	-0.289	-0.170	-0.394	-0.004	0.450
		RMSE	3.790	1.396	0.638	0.596	0.624	0.770	0.543	0.450	1.098
		MAE	2.894	1.203	0.396	0.350	0.379	0.559	0.295	0.194	0.785
	ResNetV2	R^2	**0.763**	0.060	-0.245	-0.417	-0.349	-0.149	-0.407	0.002	0.496
		RMSE	3.540	1.399	0.640	0.601	0.639	0.763	0.546	0.448	1.050
		MAE	2.759	1.235	0.402	0.356	0.398	0.564	0.298	0.193	0.751
	Xception	R^2	0.759	0.054	-0.253	-0.392	-0.306	-0.141	-0.363	0.009	0.499
		RMSE	3.567	1.404	0.642	0.595	0.629	0.761	0.537	0.447	1.047
		MAE	2.806	1.222	0.402	0.349	0.390	0.550	0.289	0.194	0.762
High + Facade	ResNet50	R^2	0.746	0.052	-0.269	-0.469	-0.328	-0.149	-0.419	-0.024	0.458
		RMSE	3.666	1.405	0.646	0.612	0.634	0.763	0.548	0.454	1.089
		MAE	2.826	1.249	0.409	0.369	0.394	0.564	0.301	0.198	0.785
	ResNetV2	R^2	**0.750**	0.031	-0.289	-0.495	-0.319	-0.146	-0.438	-0.031	0.444
		RMSE	3.636	1.421	0.651	0.617	0.632	0.762	0.552	0.455	1.103
		MAE	2.805	1.257	0.419	0.375	0.394	0.563	0.304	0.199	0.794
	Xception	R^2	0.738	0.026	-0.317	-0.474	-0.319	-0.152	-0.425	-0.070	0.426
		RMSE	3.720	1.424	0.658	0.613	0.632	0.764	0.549	0.464	1.121
		MAE	2.837	1.247	0.423	0.370	0.388	0.560	0.302	0.202	0.801
Facade	ResNet50	R^2	-0.021	-0.122	-0.557	-0.886	-0.805	-0.196	-0.438	-0.024	-0.024
		RMSE	7.346	1.528	0.715	0.693	0.739	0.779	0.552	0.454	1.497
		MAE	5.227	1.352	0.509	0.480	0.493	0.596	0.304	0.196	1.133
	ResNetV2	R^2	-0.010	-0.122	-0.557	-0.891	-0.809	-0.196	-0.438	-0.024	-0.024
		RMSE	7.305	1.528	0.715	0.694	0.740	0.779	0.552	0.454	1.497
		MAE	5.251	1.352	0.509	0.482	0.495	0.596	0.304	0.196	1.133
	Xception	R^2	-0.012	-0.122	-0.557	-0.886	-0.809	-0.196	-0.438	-0.024	-0.024
		RMSE	7.313	1.528	0.715	0.693	0.740	0.779	0.552	0.454	1.497
		MAE	5.244	1.352	0.509	0.480	0.495	0.596	0.304	0.196	1.133

Table 3: The table illustrates the results from the multi-modal benchmark employing three distinct feature extraction networks, all implemented within the same architecture as depicted in Figure 3. Additionally, this experiment incorporates a single-modal analysis using only the facade image. The highest R^2 score for each feature set test exceeding 0.6 is highlighted in bold.

counts, yield near-zero or negative values. These results indicate that when using pure facade images without additional features, the models fail to extract meaningful predictive information for interior attributes. This highlights the limitations of relying solely on facade imagery and underscores the necessity of combining image data with structured features for effective prediction.

In summary, while the addition of facade image data introduces slight improvements for specific attributes such as bedroom count, it generally leads to performance degradation for other interior attributes, particularly when the number of exterior features is reduced. The consistent underperformance for toilet and small room counts further underscores their inherent prediction difficulty.

3.2 Explainability Analysis

Figure 4 presents the heatmaps for Permutation Importance (left) and SHAP values (right), providing insights into the contributions of exterior features to the predictions of various interior attributes. From the Permutation Importance heatmap, it is clear that for interior wall length predictions, the most influential features are area, width, and house perimeter. These three features are the primary driving factors, aligning well with the results observed in the ablation study, where footprint features play a critical role. Following these, features such as facade window width, facade door width, house type (0-detached), and floor numbers emerge as secondary contributors. Features related to depth (long, short, and average) and form type (0 or 1) play relatively lesser but still notable roles. This pattern corroborates the findings from Table 2, where removing facade features leads to a performance drop, emphasising their importance in maintaining prediction accuracy.

For specific room predictions, the total room count follows a similar trend to that observed for interior wall length predictions. Key features such as floor numbers, building width, area, and facade door width are identified as dominant factors. Additionally, predictions for individual room types—such as bedrooms, lounges, dining rooms, and bathrooms—exhibit a similar dependency on the floor features. These results highlight the logical relationship between floor levels and room arrangements, which is particularly consistent with the architectural layout of UK housing.

The SHAP values heatmap, visualised using a logarithmic scale, provides a more granular representation of the contributions of each exterior feature. For interior wall length predictions, the short depth feature exhibits slightly higher importance compared to the other two depth features (long and average). Among the four attachment types, type 0 (detached) and type 2 (terraced) are identified as more significant contributors than type 1 (semi-detached) and type 3 (end-terraced). For total room count predictions, the contribution pattern closely mirrors that of interior wall length predictions but with smaller SHAP values, reflecting the relatively lower R^2 scores observed in the ablation study. In the predictions of specific rooms, floors and facade door width emerge as dominant factors, with their influence being particularly prominent in kitchen predictions. This indicates that the distribution and arrangement of different rooms across various floors are key determinants, which aligns with the typical layout structure of UK residential housing.

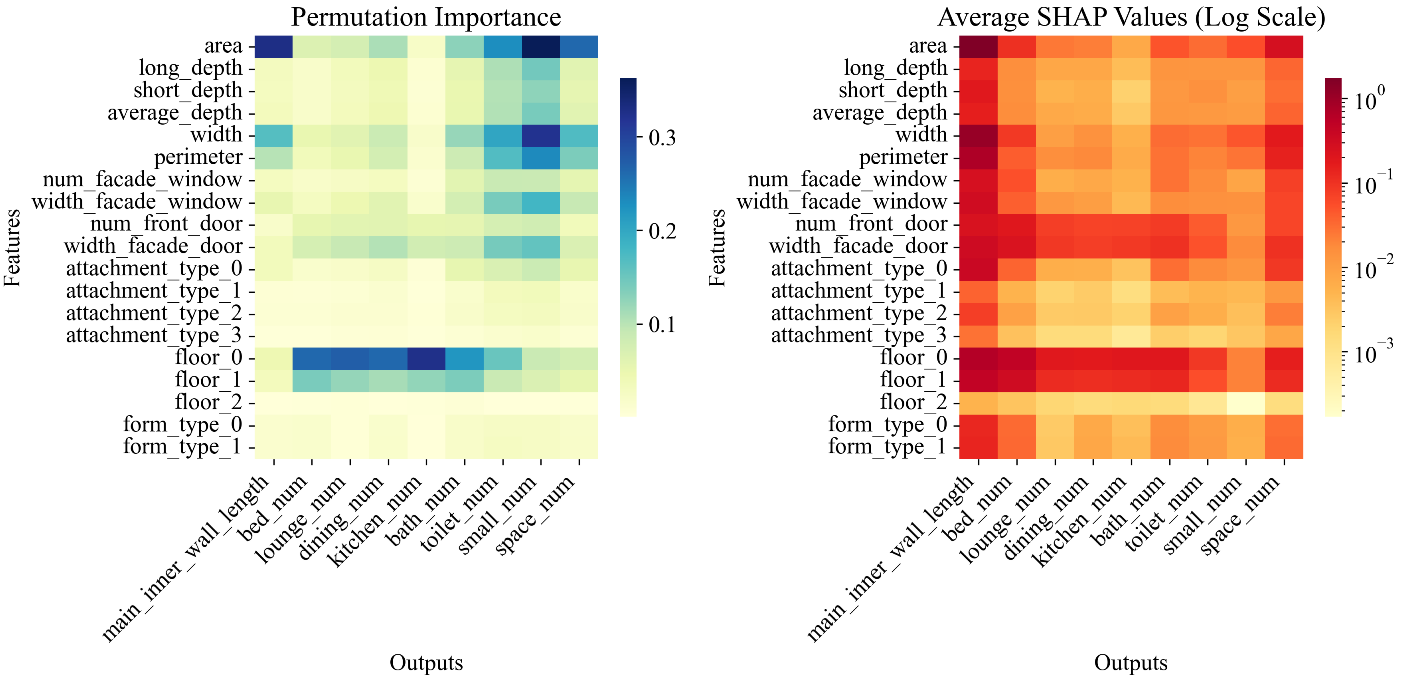

Figure 4: The figure illustrates matrices based on the permutation importance from R^2 and the average SHAP values, utilising the random forest model trained on the complete dataset. On the left is the permutation heatmap that signifies the model's dependency on particular features. On the right is the heatmap of average SHAP values, calculated by averaging the SHAP values for each individual sample. The final presentation employs a logarithmic scale to enhance readability. Raw data of drawing the figure is appended in the supporting information, SI-Fig_4.xlsx.

In summary, the heatmaps provide consistent insights into the critical exterior features that drive interior attribute predictions. Permutation Importance highlights high-level feature contributions, while SHAP values reveal a more detailed contribution structure. The results emphasise the importance of footprint features (e.g., area, width, and perimeter) and floor features for aggregate predictions like interior wall length and total room count. Additionally, facade-related features, such as window and door widths, play significant roles in specific room predictions. These findings align with the trends observed in the ablation study and reflect the logical spatial arrangements commonly seen in UK housing.

Figure 5 presents the top-5 best-performing examples (Part A) and the top-5 least-performing examples (Part B). All examples have been scaled to represent their actual sizes for consistency and comparability. In Part A, four out of the five best-performing samples correspond to first-floor layouts. These examples exhibit regular geometric sizes and well-distributed functional zones, which likely contribute to more confident and accurate predictions by the models. The clear spatial organisation and uniformity in these layouts facilitate better feature extraction and learning, resulting in higher predictive performance.

In contrast, Part B highlights the least-performing examples, all of which are ground-floor layouts. These examples lack clear functional zoning, with irregular or ambiguous spatial distributions. It is also evident in a number of the

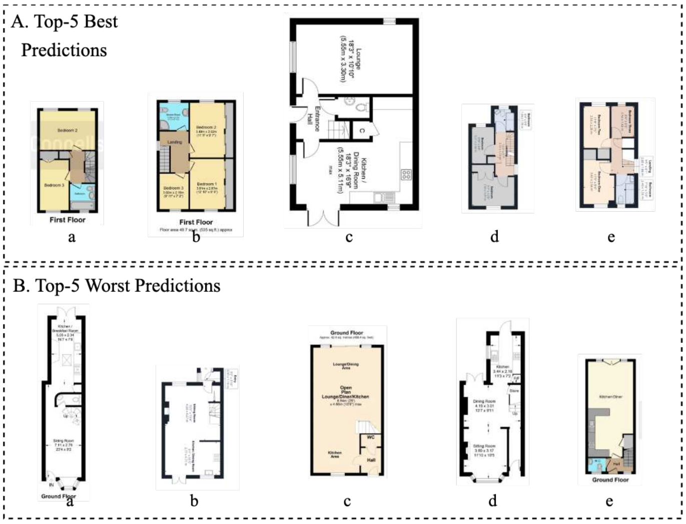

Figure 5: This figure illustrates the five most and least optimal cases by presenting their respective floor plans. All floor plans are adjusted to accurately reflect their actual dimensions.

floor plans that their internal layout has been altered during the building lifespan, e.g. walls knocked through to create an open plan space, thus the current floor plan does not match the original construction. The absence of distinct, well-defined zones introduces uncertainty in the model's predictions, leading to poor performance. Ground floors, by their nature, may also include additional structural complexities (e.g., open spaces, garages, or undefined areas), further reducing the model's ability to accurately interpret their features. In summary, the results emphasise that regular spatial layouts and well-defined functional zones are critical for achieving reliable predictions, while irregular and poorly zoned configurations, particularly on ground floors, present challenges for model performance.

4 Discussion

4.1 Scalability of Modelling House Interior Using Exterior Features

Transitioning to a circular economy and effectively managing existing building and material stocks requires a detailed understanding of the geospatial location, quantity and use of potential secondary resources. The absence of detailed information about the interior composition of buildings poses a significant challenge to achieving this,in particular through limited consideration of non-structural elements, the configuration of elements within a building, and their provided function. To address this, we develop the UKResi dataset, containing internal and external imagery (e.g. facade and room interior), geometry (e.g. building width/depth and wall length) and labelled features (e.g. room function and window/door counts) for 2,000 houses in the UK. The UKResi dataset is used to explore the relationship between interior and exterior features using a range of multi- and single- modal machine learning techniques. This reveals ability to predict internal building features (e.g. wall length and room counts) from external images and derived attributes, with potential for these attributes to be used in determining the quantity, configuration and function of internal residential material stocks across the structural, space and service layers.

The developed dataset is the first to estimate interior building structures using exterior features. Prior to this study, research on building-attribute extraction focused predominantly on remote sensing techniques to gather exterior attributes, while interior features were examined exclusively through interior data such as images or floorplans. By bridging these two domains, our dataset enables scalable interior stock accounting.

In the designed benchmark experiments, exterior features were categorised based on their accessibility and evaluated using machine learning models. The results demonstrate that interior wall length can be estimated with a high degree of accuracy using footprint-extracted features alone, achieving an average error of only 0.6 meters when compared to using a comprehensive feature set. This finding is significant as it shows that wall component dimensions can be reasonably approximated from readily available exterior data. Building on this foundation, it is plausible that the dimensions of other critical interior components, such as floor systems and roof structures, could also be estimated using similar approaches. These predictions provide an essential first step toward spatially informed material stock modelling, enabling more precise assessments of material quantities within the built environment.

Despite the promising results for wall length prediction, we observe that the accuracy of interior space predictions is highly dependent on facade features, including doors, windows, and their corresponding floor assignments. With the increasing availability of street-view imagery services and advancements in deep learning technologies, such as the Segment Anything Model (SAM) (Kirillov et al., 2023), the cost of analysing facade structures at scale has been or will foreseeably be significantly reduced. However, the scalability of this approach remains limited due to issues with data quality, particularly in regions with insufficient street-view coverage or inconsistent imagery (Hou & Biljecki, 2022). These limitations hinder the immediate large-scale application of our dataset for interior space prediction tasks.

A widely adopted solution in bottom-up stock modelling is to employ archetypes—generalised building models that represent groups of similar structures. While the scalability of direct interior space prediction remains a challenge, the developed UKResi dataset can serve as a foundation for deriving geometry-encoded archetypes through unsupervised learning techniques such as clustering. By identifying groups of buildings with similar exterior geometries that correlate to specific interior spatial configurations, these archetypes can enhance the scalability of material stock modelling efforts. This approach not only mitigates the challenges posed by limited facade data but also provides a scalable framework for estimating material stocks across larger building inventories.

In summary, while direct predictions of interior material stocks face data-related scalability challenges, the demonstrated feasibility of wall length estimation highlights the potential for similar approaches to estimate other interior components. The development of geometry-informed archetypes offers a practical pathway to scale these methods, supporting broader applications in circular economy practices, such as material stock accounting and resource recovery.

4.2 Contributions to Digital Twins of Cities

The developed UKResi dataset systematically captures interior wall and room function information, making it a valuable resource for reconstructing indoor spaces. In addition to the features used for benchmark tests, the dataset includes linkage information such as interior doors and room connectivity, enabling the creation of a room connection graph based on room types and spatial relationships. For example, consider a ground floor with a lounge, bathroom, and kitchen, where the lounge is accessible from both the bathroom and kitchen. This layout can be represented as a graph in which each room is a node, and edges indicate the presence of a door between two rooms. By concluding the linkage patterns, these graphs can serve as a foundational representation of interior layouts, providing a structured framework for understanding indoor spatial compositions. This has potential applications in assessing the building-level circular economy potential of existing stocks, including the potential for adaptation through subdivision (i.e. splitting one property into multiple), conversion (i.e. changing of the use of a property or room) and/or extension (i.e. adding new space above or adjacent to a property).

Leveraging advancements in generative artificial intelligence (AI) technologies, such as the *HouseGAN* series models (Nauata et al., 2020, 2021), these connection graphs can be used to generate interior floor layouts. Specifically, when combined with building footprint constraints, the generative models can produce pseudo-interior layouts that approximate realistic indoor configurations. This process enables the reconstruction of indoor spaces, even in the absence of direct indoor data, addressing one of the key challenges in creating digital twins of cities. The significance of this lies in the role of digital twins as dynamic, virtual representations of urban environments. While outdoor and exterior building information can be readily obtained through remote sensing and street-view imagery, acquiring detailed indoor data remains a significant hurdle, particularly at scale. By utilising the UKResi dataset and the predictive capabilities of AI-driven models, it becomes feasible to reconstruct indoor spatial layouts, thereby bridging the data gap for indoor environments. This reconstructed spatial data can significantly increase the level of detail (LoD) of urban digital twins from LoDs 2 or 3 (3D models without and with external architectural details, respectively) to LoD4, in which interior details are included (Jeddoub et al., 2023). Such inclusion of interior elements is key to enable a variety of sustainability-related modelling such as energy modelling, occupancy simulations, retrofit planning, and material stock assessments.

In summary, the UKResi dataset, with its rich feature set and connectivity information, provides a robust foundation for reconstructing indoor spaces. Combined with generative AI models and footprint constraints, it offers a scalable solution for approximating indoor layouts. This approach holds significant potential for advancing digital twins of cities, where comprehensive indoor data is critical for enabling more accurate simulations, sustainable urban planning, and smarter building management.

4.3 Future Work

While the UKResi dataset provides a solid foundation for linking exterior features to interior spatial conditions, several areas of improvement and expansion remain to enhance its utility and scalability further. Firstly, future work will focus on developing more detailed archetypes that incorporate a wider range of exterior features for improved material stock modelling and room-level predictions. By leveraging advanced unsupervised learning techniques such as clustering, we aim to create exterior-encoded archetypes that can represent diverse building typologies more accurately. These archetypes will help generalise the relationship between exterior and interior features across similar buildings, thus supporting large-scale bottom-up stock modelling and resource assessments.

Secondly, the dataset will be expanded to include a greater variety of house types to address limitations observed in predicting ground-level interiors. The current performance drop for ground-floor predictions, particularly for attributes such as room counts and spatial layouts, can be attributed to the inherent irregularity and variability (particularly over time due to renovations) in these configurations. By incorporating houses from different architectural styles, construction periods, and geographical regions, the dataset will better represent the diversity of residential buildings. This expanded coverage will help mitigate the challenges associated with ground-level predictions, ensuring the models achieve more robust and reliable performance. These advancements will further support the creation of accurate and scalable digital twins of cities, facilitating sustainable urban planning and resource management.

Author Contributions

Menglin Dai: Conceptualisation; Formal analysis; Methodology; Software; Validation; Visualisation; Writing - original draft; Writing - review & editing. **Jakub Jurczyk**: Data curation; Visualisation; Writing - review & editing. **Charles Gillott**: Conceptualisation; Writing - original draft; Writing - review & editing. **Kun Sun**: Visualisation; Writing - review & editing. **Maud Lanau**: Writing - review & editing. **Gang Liu**: Conceptualisation; Funding acquisition; Project administration; Resources; Supervision; Writing - review & editing. **Danielle Densley Tingley**: Conceptualisation; Funding acquisition; Project administration; Resources; Supervision; Writing - review & editing.

Data and Code Availability Statement

The code used in the benchmark experiment will be made available on the designated GitHub repository: https://github.com/MerlinDai/UKResi. The developed UKResi dataset will be available upon request from the corresponding author upon reasonable request.

Supporting Information

1. SI-Data_Distributions.doc This supporting information provides additional statistical and spatial visualisations of the constructed UKResi dataset.
2. SI-fig_4.xlsx This supporting information provides the values to generate Figure 4.
3. SI-attributes_summary.xlsx This supporting information provides the attributes definitions of the constructed UKResi dataset.

Funding

We appreciate financial support from the National Natural Science Foundation of China (71991484), EPSRC BuildZero: transforming the UK's buildings for zero material extraction, zero carbon and zero waste, United Kingdom [EP/Y530578/1], EPSRC Multi-Scale, Circular Economic Potential of Non-Residential Building Scale [EP/S029273/1], and the Fundamental Research Funds for the Central Universities of Peking University.

References

Abadi, M. (2016). TensorFlow: learning functions at scale. *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming*, 1–1.

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347.

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., & Weaver, J. (2010). Google Street View: Capturing the World at Street Level. *Computer*, *43*(6), 32–38.

Arceo, A., MacLean, H. L., & Saxe, S. (2023). Material intensity in single-family dwellings: Variability between locations, functional unit and drivers of material use in toronto, perth, and luzon. *Resources, Conservation and Recycling*, *188*, 106683.

Arceo, A., Tham, M., Guven, G., MacLean, H. L., & Saxe, S. (2021). Capturing variability in material intensity of single-family dwellings: A case study of toronto, canada. *Resources, Conservation and Recycling*, *175*, 105885.

Arora, M., Raspall, F., Cheah, L., & Silva, A. (2019). Residential building material stocks and component-level circularity: The case of singapore. *Journal of Cleaner Production*, *216*, 239–248.

Brand, S. (1995). *How buildings learn: What happens after they're built*. Penguin.

Budroni, A., & Boehm, J. (2010). Automated 3d reconstruction of interiors from point clouds. *International Journal of Architectural Computing*, *8*(1), 55–73.

Buyukdemircioglu, M., Can, R., Kocaman, S., & Kada, M. (2022). Deep learning based building footprint extraction from very high resolution true orthophotos and ndsm. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, *2*, 211–218.

Cai, B., Shao, Z., Huang, X., Zhou, X., & Fang, S. (2023). Deep learning-based building height mapping using sentinel-1 and sentienl-2 data. *International Journal of Applied Earth Observation and Geoinformation*, *122*, 103399.

Cao, Y., & Weng, Q. (2024). A deep learning-based super-resolution method for building height estimation at 2.5 m spatial resolution in the northern hemisphere. *Remote Sensing of Environment*, *310*, 114241.

Cao, Z., Myers, R. J., Lupton, R. C., Duan, H., Sacchi, R., Zhou, N., Reed Miller, T., Cullen, J. M., Ge, Q., & Liu, G. (2020). The sponge effect and carbon emission mitigation potentials of the global cement cycle. *Nature communications*, *11*(1), 3777.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 833–851.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251–1258.

Dai, M. (2023). *Characterising english residential housing using street view capture and deep learning techniques* [Doctoral dissertation, University of Sheffield].

Dai, M., Jurczyk, J., Arbabi, H., Mao, R., Ward, W., Mayfield, M., Liu, G., & Tingley, D. D. (2024). Component-level residential building material stock characterization using computer vision techniques. *Environmental Science & Technology*, *58*(7), 3224–3234.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, 248–255.

Department for Communities and Local Government. (2016). English housing survey 2014 to 2015: Housing stock report [Accessed: 2024-12-16]. https://www.gov.uk/government/statistics/english-housing-survey-2014-to-2015-housing-stock-report

Di Donato, M., Lomas, P. L., & Carpintero, Ó. (2015). Metabolism and environmental impacts of household consumption: A review on the assessment, methodology, and drivers. *Journal of Industrial Ecology*, *19*(5), 904–916.

Fishman, T., Mastrucci, A., Peled, Y., Saxe, S., & van Ruijven, B. (2024). RASMI: Global ranges of building material intensities differentiated by region, structure, and function. *Scientific Data*, *11*(1), 418.

Frantz, D., Schug, F., Wiedenhofer, D., Baumgart, A., Virág, D., Cooper, S., Gómez-Medina, C., Lehmann, F., Udelhoven, T., van der Linden, S., et al. (2023). Unveiling patterns in human dominated landscapes through mapping the mass of us built structures. *Nature Communications*, *14*(1), 8014.

Google Research. (2021). Open Buildings: A Dataset for Mapping Buildings at Scale [https://sites.research.google/gr/open-buildings/ (Accessed: 2024-06-17)]. https://sites.research.google/gr/open-buildings/

Guo, H., Du, B., Zhang, L., & Su, X. (2022). A coarse-to-fine boundary refinement network for building footprint extraction from remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *183*, 240–252.

Haberl, H., Wiedenhofer, D., Schug, F., Frantz, D., Virág, D., Plutzar, C., Gruhler, K., Lederer, J., Schiller, G., Fishman, T., et al. (2021). High-resolution maps of material stocks in buildings and infrastructures in austria and germany. *Environmental science & technology*, *55*(5), 3368–3379.

Hancock, H. (2022). Rightmove vs Zoopla vs OnTheMarket: The Ultimate Visual Guide to Property Portals in the UK. *Online Marketplaces*. https://www.onlinemarketplaces.com/articles/ultimate-visual-guide-to-property-portals-in-the-uk/

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Heeren, N., Mutel, C. L., Steubing, B., Ostermeyer, Y., Wallbaum, H., & Hellweg, S. (2015). Environmental impact of buildings: What matters? *Environmental science & technology*, *49*(16), 9832–9841.

Hou, Y., & Biljecki, F. (2022). A comprehensive framework for evaluating the quality of street view imagery. *International Journal of Applied Earth Observation and Geoinformation*, *115*, 103094.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7132–7141.

Hu, Q., Yang, B., Khalid, S., Xiao, W., Trigoni, N., & Markham, A. (2021). Towards semantic segmentation of urban-scale 3d point clouds: A dataset, benchmarks and challenges. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4977–4987.

Huang, W., Olson, A. W., Khalil, E. B., & Saxe, S. (2024). Image-based prediction of residential building attributes with deep learning. *Journal of Industrial Ecology*.

Jeddoub, I., Nys, G.-A., Hajji, R., & Billen, R. (2023). Digital twins for cities: Analyzing the gap between concepts and current implementations with a specific focus on data integration. *International Journal of applied earth observation and geoinformation*, *122*, 103440.

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. (2023). Segment anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4015–4026.

Kissinger, M., & Damari, Y. (2021). Household metabolism: Integrating socio-economic characteristics and lifestyles on individual and national scales as a mean for advancing environmental management. *Journal of Environmental Management*, *279*, 111526.

Krapf, S., Mayer, K., & Fischer, M. (2023). Points for energy renovation (pointer): A point cloud dataset of a million buildings linked to energy features. *Scientific Data*, *10*(1), 639.

Lanau, M., & Liu, G. (2020). Developing an urban resource cadaster for circular economy: A case of odense, denmark. *Environmental science & technology*, *54*(7), 4675–4685.

Lanau, M., Liu, G., Kral, U., Wiedenhofer, D., Keijzer, E., Yu, C., & Ehlert, C. (2019). Taking stock of built environment stock studies: Progress and prospects. *Environmental science & technology*, *53*(15), 8499–8515.

Li, H., Yuan, Z., Dax, G., Kong, G., Fan, H., Zipf, A., & Werner, M. (2023). Semi-supervised learning from street-view images and openstreetmap for automatic building height estimation. *arXiv preprint arXiv:2307.02574*.

Li, X., Song, L., Liu, Q., Ouyang, X., Mao, T., Lu, H., Liu, L., Liu, X., Chen, W., & Liu, G. (2023). Product, building, and infrastructure material stocks dataset for 337 chinese cities between 1978 and 2020. *Scientific Data*, *10*(1), 228.

Lin, M. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.

Liu, J., Wang, M., Zhang, C., Yang, M., & Li, Y. (2020). Material flows and in-use stocks of durable goods in chinese urban household sector. *Resources, Conservation and Recycling*, *158*, 104758.

Lundberg, S. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Mahmoud, M., Chen, W., Yang, Y., & Li, Y. (2024). Automated bim generation for large-scale indoor complex environments based on deep learning. *Automation in Construction*, *162*, 105376.

Miatto, A., Fasanella, Y., Mainardi, M., & Borin, P. (2023). Correlation between building size and material intensity in residential buildings. *Resources, Conservation and Recycling*, *197*, 107093.

Microsoft. (2024). Microsoft Global ML Building Footprints [https://github.com/microsoft/GlobalMLBuildingFootprints (Accessed: 2024-06-17)]. https://github.com/microsoft/GlobalMLBuildingFootprints

Milojevic-Dupont, N., Wagner, F., Nachtigall, F., Hu, J., Brüser, G. B., Zumwald, M., Biljecki, F., Heeren, N., Kaack, L. H., Pichler, P.-P., et al. (2023). Eubucco v0. 1: European building stock characteristics in a common and open database for 200+ million individual buildings. *Scientific Data*, *10*(1), 147.

Ministry of Housing, Communities & Local Government. (2010). English housing survey 2008: Housing stock report [https://www.gov.uk/government/statistics/english-housing-survey-2008-housing-stock-report (Accessed: 2024-12-16)]. https://www.gov.uk/government/statistics/english-housing-survey-2008-housing-stock-report

Ministry of Housing, Communities & Local Government. (2018). English Housing Survey, Households Report, 2017-18 [https://www.gov.uk/government/statistics/english-housing-survey-2017-to-2018-households (accessed 2024-06-06)].

Nasiri, B., Kaasalainen, T., & Hughes, M. (2023). Estimating the material intensity of wooden residential houses in finland. *Resources, Conservation and Recycling*, *198*, 107142.

Nauata, N., Chang, K.-H., Cheng, C.-Y., Mori, G., & Furukawa, Y. (2020). House-gan: Relational generative adversarial networks for graph-constrained house layout generation. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, 162–177.

Nauata, N., Hosseini, S., Chang, K.-H., Chu, H., Cheng, C.-Y., & Furukawa, Y. (2021). House-gan++: Generative adversarial layout refinement network towards intelligent computational agent for professional architects. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13632–13641.

Office for National Statistics. (2023). Housing, England and Wales: Census 2021 [https://www.ons.gov.uk/peoplepopulationandcommunity/housing/bulletins/housingenglandandwales/census2021 (accessed 2024-06-06)].

Ordnance Survey. (2023). Mastermap topography layer [https://www.ordnancesurvey.co.uk/business-government/products/mastermap-topography (accessed 2024-05-15)].

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. *the Journal of Machine Learning research*, *12*, 2825–2830.

Pei, W., Biljecki, F., & Stouffs, R. (2024). Techniques and tools for integrating building material stock analysis and life cycle assessment at the urban scale: A systematic literature review. *Building and Environment*, 111741.

Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. *IEEE signal processing magazine*, *34*(6), 96–108.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. *Proceedings of the AAAI conference on artificial intelligence*, *31*.

UK Government. (2024). Our plan to build more homes [https://www.gov.uk/government/news/our-plan-to-build-more-homes (Accessed: 2024-06-17)]. https://www.gov.uk/government/news/our-plan-to-build-more-homes

United Nations. (2015). Sustainable development goals [https://www.un.org/sustainabledevelopment/sustainable-development-goals/ (accessed 2023-06-06)].

United Nations Framework Convention on Climate Change (UNFCCC). (2015). Paris agreement [https://unfccc.int/sites/default/files/english_paris_agreement.pdf (Accessed: 2024-06-17)]. https://unfccc.int/sites/default/files/english_paris_agreement.pdf

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. *Proceedings of the European conference on computer vision (ECCV)*, 3–19.

Xu, P., Zhu, X., & Clifton, D. A. (2023). Multimodal learning with transformers: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(10), 12113–12132.

Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. *IEEE transactions on knowledge and data engineering*, *34*(12), 5586–5609.

Zhong, X., Deetman, S., Tukker, A., & Behrens, P. (2022). Increasing material efficiencies of buildings to address the global sand crisis. *Nature Sustainability*, *5*(5), 389–392.

Zhong, X., Hu, M., Deetman, S., Steubing, B., Lin, H. X., Hernandez, G. A., Harpprecht, C., Zhang, C., Tukker, A., & Behrens, P. (2021). Global greenhouse gas emissions from residential and commercial building materials and mitigation strategies to 2060. *Nature Communications*, *12*(1), 6126.

Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., & Torralba, A. (2019). Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision*, *127*, 302–321.