Measuring Residential House Interior Wall Lengths Based on External Features Using Machine Learning Techniques

Jakub Jurczyk
Email: jjurczyk1@sheffield.ac.uk
Department of Civil and Structural Engineering, Faculty of Engineering
Project Supervisors: Dr Menglin Dai, Dr Danielle Densley-Tingley

Introduction

Understanding the energy performance of buildings and formulating effective renovation strategies necessitates a thorough comprehension of building material stock. However, the limitations associated with manual analysis, primarily its time-consuming nature and inadequacies, deem it not suitable for large scale assessment. To address this, an automated approach has been proposed. This study aims to present an innovative method for automatically assessing the internal wall length—a pivotal factor in accurately calculating the complete material composition of a house.

Existing Research

Publically available datasets:

- **CVC-FP** [1]: Offers 122 samples for training models on automatic wall segmentation and room labelling within floor plans, revealing insights into architectural elements' relationships.
- **FPLAN-POLY** [2]: Contains 42 vectorized real floor plan images, serving the purpose of training symbol spotting models.
- **ROBIN** [3]: Consists of 510 floorplans, aiming to train an object analysis model that matches building layouts to user needs.
- **SESYD** [4]: A database of 1000 synthetically generated floor plan samples, developed for symbol spotting and identifying architectural objects and furniture.
- **BRIDGE** [5]: Comprises 13,000 floor plan samples used to generate text-based descriptions for existing floor plans.

Despite the valuable contributions of the mentioned datasets to the field, they lack vital elements necessary for our research, specifically pertaining to English and Welsh housing. They notably lack information about the location of houses and do not include exterior photos. These deficiencies render them inadequate for effectively training our model tailored to our research objectives. Given these constraints and the fact that existing datasets were primarily created and labelled for computer vision algorithms.

Data Collection

The data about houses was obtained by studying floor plans and gathering information from British real estate online services (Zoopla and Savills). A set 300 houses, including detached, semi-detached, and terraced types from various regions of England and Wales, was collected to ensure dataset representativeness.

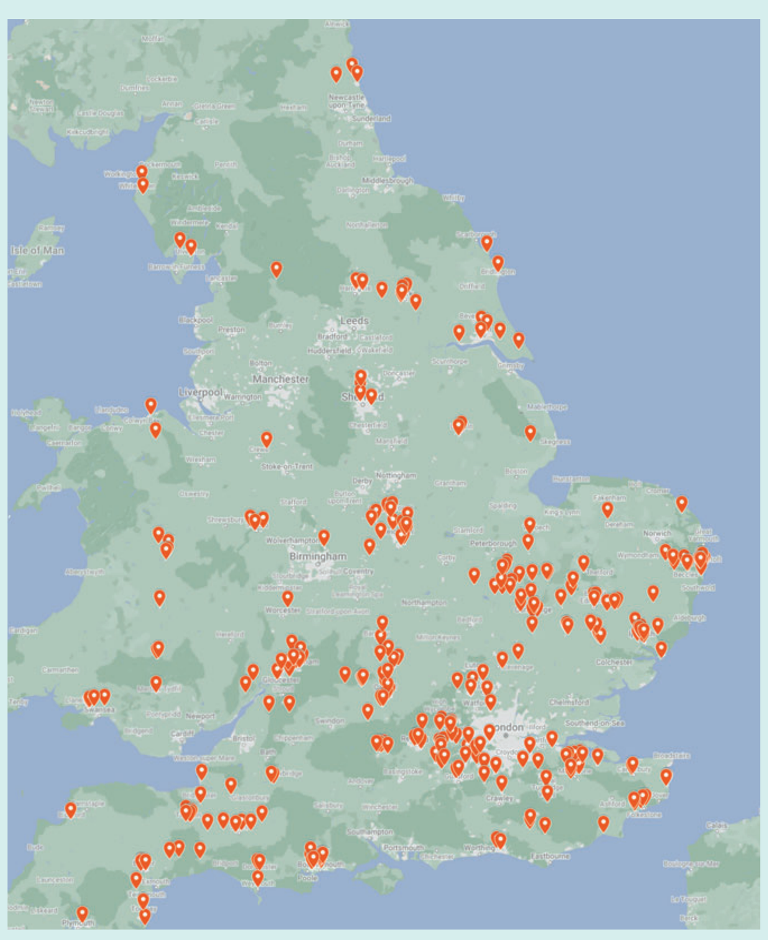


Figure 1. Location of the sampled houses.

Data Analysis

In the dataset, every floor plan is treated as a separate sample that need to be analysed. The following information is being extracted from each floor plan to train the algorithm:

Classification Variables → 1. Building Type; 2. Floor Number

Floorplan Footprint → 3. Depth of the House; 4. Width of the House; 5. Perimeter of the House

Opening Information → 6. Number of facade openings; 7. Width of the facade openings

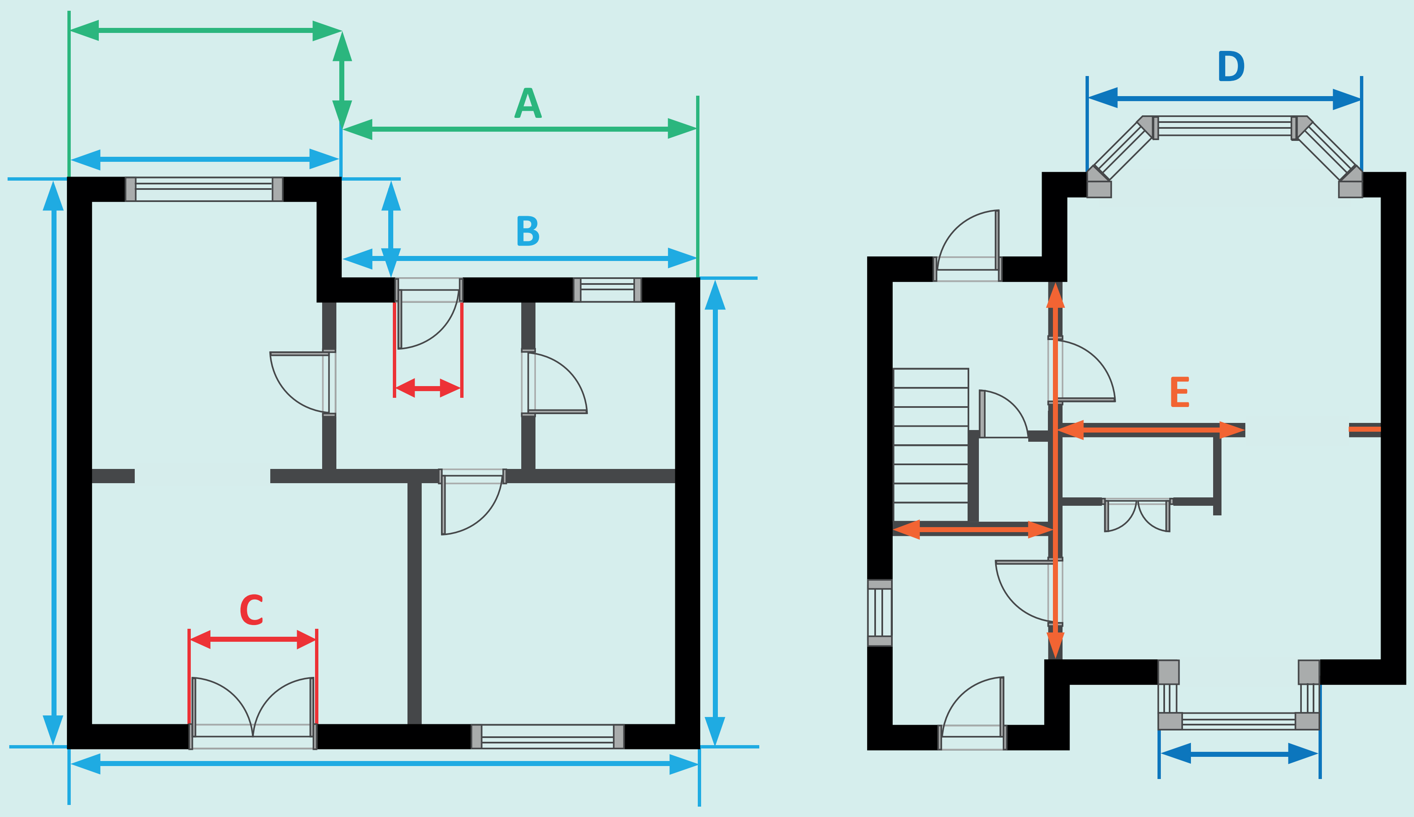
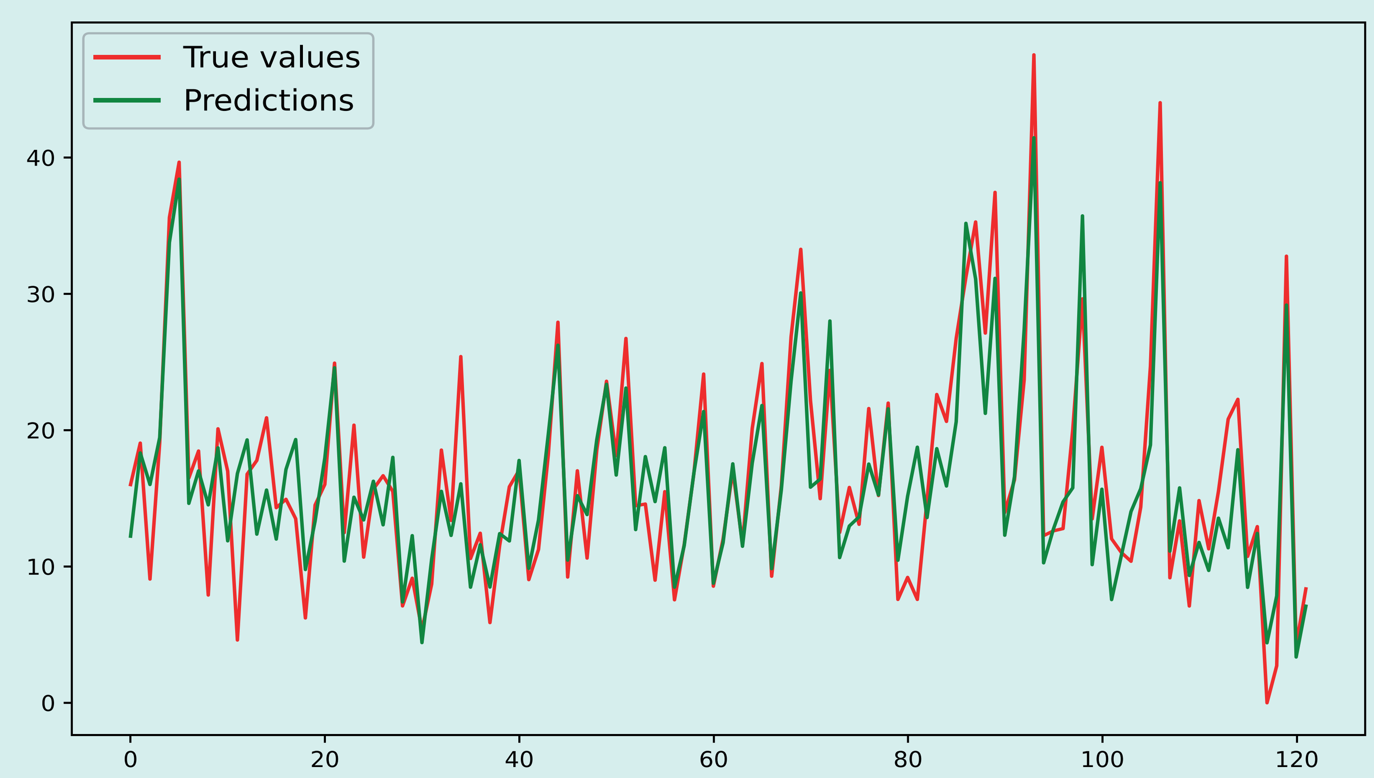


Figure 2. Examples of sampled parameters. A – front parameter, B – total parameter, C – door width, D – window opening, E – internal wall length

Model Training and Results

- After testing different regression algorithms, SVM (Supervised Vector Machine) gave the most accurate predictions.
- Average R^2 value given by 5-fold test is equal to **70%**. In other words, **predicted values match the true values in 70%**).



Graph 1. Visual comparison of the predicted and true values.

Discussion

Some floor plan types are more represented in the dataset, causing the algorithm to be biased towards them, impacting prediction accuracy. Expanding the dataset and ensuring the variety of the samples would enhance prediction accuracy.

Bibliography

1. de las Heras, L.-P. et al. (2015) 'CVC-FP and Sgt: A new database for structural floor plan analysis and its groundtruthing tool', International Journal on Document Analysis and Recognition (IJDAR), 18(1), pp. 15–30. doi:10.1007/s10032-014-0236-5.
2. Marc¸al Rusinol, Agn ˜ es Borr ´ as, and Josep Llad ` os. Relational ´indexing of vectorial primitives for symbol spotting in linedrawing images. Pattern Recognition Letters, 31(3):188–
3. D. Sharma, N. Gupta, C. Chattopadhyay and S. Mehta, "DANIEL: A Deep Architecture for Automatic Analysis and Retrieval of Building Floor Plans," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017, pp. 420-425, doi: 10.1109/ICDAR.2017.76
4. Delalandre, M.; Valveny, E.; Pridmore, T.; Karatzas, D. Generation of synthetic documents for performance evaluation of symbol recognition & spotting systems.Int. J. Doc. Anal. Recognit. (IJDAR)2010,13, 187–207.
5. Shreya Goyal, Vishesh Mistry, Chiranjoy Chattopadhyay, and Gaurav Bhatnagar. Bridge: Building plan repository for image de scription generation, and evaluation. In 2019 International Conference on Document Analysis and Recognition (ICDAR), pages 1071–1076. IEEE, 2019.