IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

Deep multimodal learning for residential building energy prediction

To cite this article: Y Sheng *et al* 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1078** 012038

View the article online for updates and enhancements.

Deep multimodal learning for residential building energy prediction

Y Sheng[1], W OC Ward[1], H Arbabi[1], M Álvarez[2] and M Mayfield[1]

[1]Department of Civil and Structural Engineering, The University of Sheffield, UK
[2]Department of Computer Science, University of Manchester, Manchester, UK

E-mail: ysheng8@sheffield.ac.uk

Abstract. The residential sector has become the second-largest energy consumer since 1987 in the UK. Approximately 24 million existing dwellings in England made up over 32% of the overall energy consumption in 2020. A robust understanding of existing buildings' energy performance is therefore critical in guiding proper home retrofit measures to accelerate towards meeting the UK's climate targets. A substantial number of predictions at a city scale rely on available data, e.g., Energy Performance Certificates (EPCs) and GIS products, to develop statistical and machine learning models to estimate energy consumption. However, issues with existing data are not negligible. This work adopted the idea of deep multimodal learning to study the potential for using Google Street View (GSV) images as an additional input for residential building energy prediction. 20,031 GSV images of 5,933 residential buildings in central Barnsley, UK, have been selected for a case study. All images were pre-processed using a state-of-the-art object detection algorithm to minimise the noise caused by other elements that may appear nearby. Building specifications that cannot be easily determined by the appearance are extracted from existing EPC information as text-based inputs for prediction. A multimodal model was designed to jointly take images and texts as inputs. These inputs are first propagated through a convolutional neural network and multi-layer perceptron, respectively, before being combined into a connected network for final energy prediction. The multi-input model was trained and tested on the case study area and predicted an annual energy consumption with a mean absolute difference of 0.01kWh/m^2 per annum on average compared with what is recorded in the EPC. The difference between the predicted results and the EPC may also provide some hints on the bias the certificates potentially contain.

Key Words: Residential building energy; Deep multimodal learning; EPC; Google Street View

1. Introduction

The importance of retrofitting existing housing stocks to a better standard has gradually been recognised by governments worldwide as one of the important mechanisms to tackle climate change. Since 2020, a large majority of the world's population has been forced to work from home due to the widespread virus COVID-19. In the UK, the official statistics suggested that the residential sector has become the only sector that witnessed an increase in energy consumption in 2020 [1]. The Department for Business, Energy and Industrial Strategy (BEIS) in the UK is investing almost £6.6 billion to support more retrofit projects [2]. To efficiently implement retrofit projects, a proper understanding of the energy performance of existing housing stocks is important. This includes understanding the properties' current insulation conditions and their corresponding energy consumption.

A large number of existing studies have been conducted to estimate building energy consumption. Studies at a large scale usually adopt a data-driven approach and apply a machine learning-based model to estimate the potential energy demand [3,4]. These studies rely on available data that provides a range of attributes that affect buildings' energy performance, including the type of building, the fabric of the built envelope, and the insulation conditions [4,5]. Various datasets are developed over the years for this purpose, but limitations exist. For example, Energy Performance Certificate (EPC) is an official document that was first introduced in 2002 by the European Union to improve energy performance and raise public awareness [4,6]. The certificate contains information on building characteristics assessed by qualified inspectors. From these characteristics, the energy consumption is calculated to represent the energy usage under standard operational conditions. This consumption is used by the EPC as one of the criteria for energy rating classification and guidance for potential retrofit measures [4,6]. However, studies show the EPC recorded consumption exhibits large discrepancies with actual usage [7].

Researchers are attempting to introduce alternative datasets and methodologies to replace the role EPCs play in providing housing characteristics, for instance, real-estate evaluation reports. The real-estate report is a popular dataset in computer vision but has only been introduced into energy estimation recently [8,9]. Similarly, Google Street View (GSV) images are popular dataset in neighbourhood and traffic analysis [10,11]. The GSV sensing vehicles drive around at the same time recording images of buildings alongside the roads, so this study proposes that this type of data should have the capability to provide property visual characteristics for energy performance studies.

Most existing studies only use a single data source as input [8,9,12]. This study proposes a multimodal deep learning network that multiple datasets are used at the same time to predict building energy consumption. The concept of using multimodal for deep learning tasks has been gradually attracting more attention over the last decade, especially in the field of expression recognition and medical diagnosis [13,14]. This study proposes that the visual representations of buildings extracted from the GSV can be adopted, combined with textual information from the EPCs, to construct a multimodal deep learning network. The two modalities can be used to predict the energy consumption individually, at the same time to provide additional explanations and cross-validation to each other towards more efficient and more accurate predictions.

The remainder of this study is structured in five sections. Section 2 provides a brief literature review of the theoretical background of building energy performance analysis and related work, leading to the motivation of developing a multimodal deep learning network for energy prediction. Section 3 explains the methodology of the developed algorithm. The results are presented and discussed in section 4, and finally, conclude in section 5.

2. Literature review

2.1. Existing residential building energy study

The existing approaches can be divided into three subgroups by the main methodologies used: data-driven; physics-based; and hybrid approaches. Studies using data-driven approaches usually apply machine learning algorithms to building characteristics and historical consumption data for city-scale building energy prediction [3]. When a historical dataset is not available, for instance, for newly developed constructions, physics-based approaches can be adopted. However, physics-based approaches usually require professional knowledge about theories of heat transfer [4]. It also involves a large amount of time and uncertainties in data preparation, parameter assumption and model setting, so is usually applied to individual objects [3]. Both data-driven and physics-based methods have their limitations, thereby a hybrid approach was introduced to combine both techniques, for instance, using data-driven algorithms to prepare data used in the physics-based approach [3,15].

Because this research aims at studying building energy consumption at a regional level, a data-driven approach is adopted and will be the focus of the rest of the literature review. Two main findings verified the capability of such a method in estimating residential building energy consumption [5]:

a. buildings constructed in the same place and time tend to have similar building characteristics; and
b. buildings with similar characteristics tend to have similar energy needs.

These two findings also provide hints on the features that have close associations with building energy performance. These features can be grouped into geometric features (such as the type of building, and the footprint area) and thermal features (for example, the insulation conditions). A substantial amount of existing literature has investigated and experimented with various data inputs and reviewed their limitations.

2.1.1. *Issues with existing data for data-driven approach*. The nature of the data-driven approach suggests a relatively high dependency on the accuracy of available data. Existing literature has used data from a range of sources including GIS products [16,17], EPCs [7,12], real estate evaluation reports [18], and also data collected in-situ with specialised sensors [19].

In-situ collection of thermal data usually involves infrared thermography, a non-destructive technique for object auditing and material characterisation. Collecting the data using thermal sensors requires pre-hand knowledge of weather conditions, u-value of the material and access to the internal area to calculate the temperature difference [20], which is not applicable for this study.

GIS products refer to a database that uses or involves a GIS platform to extract, store and exchange knowledge of building characteristics. One of the most used data is the 3D CityGML. It is an open data model that aims to provide a standard mechanism to describe and store building models in five different levels of detail [21]. However, due to the lack of completeness, most of the existing studies are only conducted using the Level of Detail 1 (LoD1) model [17,22]. In LoD1, buildings are represented by basic geometric floor plans with extruded elevations referring to the building heights. With only the area size and building height, the bias of LoD1 is not negligible as the data is missing a great number of building elements that play important roles in scaling the buildings' energy efficiency, such as the building fabric and insulation conditions [5].

EPCs to some extent contain the information LoD1 neglected. EPC contains information describing the shape, type, and insulation conditions of the assessed property. This instrument is not compulsory in Europe, so its utilisation varies across different countries, but it is one of the compulsory documents in the UK for a property to be able to sell or rent [4,23,24]. The Building Research Establishment's (BRE) Standard Assessment Procedure uses EPC information to calculate and classify buildings with different energy efficiencies into 7 rating classes [25]. The EPCs in the UK are publicly available and have been a popular dataset for UK housing stock studies. However, more studies are taking place to examine the quality of the EPCs and the potential gaps between the EPC-estimated energy demand and the actual consumption. Crawley et al. [26] examined the existing valid EPCs and discovered that 1.6 million buildings in England and Wales are having multiple EPCs of different ratings. Burman, Mumovic and Kimpian [27] studied a secondary school in Northwest England and found that there is a 3.1% of difference per year between EPC suggested energy consumption and the building's actual usage.

The real-estate evaluation report is a popular source of data in the field of computer vision [9,18], but relatively new in building energy prediction. Researchers consider the photos taken by the real estate are good visual representations showing rich information of the building and its surroundings [9]. Despotovic et al. [18] developed a workflow to extract building element images (e.g., doors, roofs, windows) to classify the corresponding housing energy ratings. They applied the network to 2,065 different houses and achieved a classification accuracy of 62% when predicting the energy rating. This case study shows the potential of applying building photographs in housing energy prediction. But it has similar disadvantages as the CityGML LoD1 database has that the images provide more descriptions of the properties' physical characteristics but are limited in actual insulation conditions.

2.1.2. *Google street view images as a source of data*. Case studies using images from real estate evaluation reports show the potential of using photographs for building energy prediction, but it is not publicly available. Google street view images (GSVs) can be a potential alternative source of image data for building energy prediction. Similar to real estate reports, GSV has been widely used in computer vision studies. Example studies have been conducted with GSVs to derive neighbourhood socio-demographic patterns [28] and traffic auditing [10]. When the GSV van drives around the city, the captured scenes are publicly available and can be downloaded with API. Yuan J and Cheriyadat AM [29] conducted building height and facade estimation using OSM building footprint and GSV. They proposed a methodology to estimate camera position and project street scenes to 2D maps. This study shows that GSV images can be used as descriptions for building characteristics, such as building height and facade material, which is critical in building energy estimation.

2.2. Deep multimodal learning

The lack of accurate data is, however, not unique to building energy analysis. There is a growing interest in the potential of developing deep learning models that use multiple inputs [13,14], named deep multimodal learning. The nature of this approach is to accept heterogeneous cues from different modalities for additional and potentially more comprehensive knowledge of a given task. Similar to the five senses human beings have, data can be classified under different modalities, such as image (visual), text (word), audio (sound), and physiological signals. Existing applications are in the field of face recognition, medical diagnosis, and self-driving systems [13,14], but few applications are in building energy prediction. This study considers that describing a property using different modalities with a multimodal deep learning approach may to some extent reduce the level of bias compared with unimodal approaches where only one type of data is used.

Table 1. Selected building energy performance-related information from EPC

Type	Data	Description
Reference	Building reference number	A unique number assigned to each EPC
	Full address	The full property address
Numeric	Total floor area	The total floor area in m^2
	Habitable rooms	Number of habitable rooms
	Heated rooms	Number of rooms that can be heated
	Lighting description*	Percentage of low energy lighting
Categorical	Property type	Type of property (e.g., house)
	Built form	Building type (e.g., detached)
	Floor description*	Type of floor and insulation condition (e.g., solid, no insulation)
	Window description*	Type of window and glazing (e.g., double glazing)
	Wall description*	Type of wall and insulation conditions (e.g., filled cavity)
	Roof description*	Type of roof and insulation conditions (e.g., pitched)
	Main heating*	Type of main heating used (e.g., boiler and radiator)
	Main fuel*	Type of main fuel used for central heating (e.g., mains gas)
Output	Energy consumption	The current energy consumption recorded by the EPC in kWh/m^2 per year

3. **Methodology**

The overall workflow of this study will be explicitly explained in the following subsections. This multimodal study mainly used two datasets: EPC certificates and GSV images. The workflow can be divided into three main stages: the data acquisition; the pre-processing of the raw data, and finally the development and training of the deep multimodal learning network.

3.1. Data acquisition and pre-processing

3.1.1. *Energy performance certificates*. The UK government has established a website for EPC records that allows access and download. The downloaded information is filtered, and only the variables that are most related to the current housing energy performance are used for prediction [5] (Table 1).

After the preliminary study of the EPC records, the issues other studies encountered are observed [26], that the records contain abnormal entries and multiple records for one property. The first step is to filter these data. Abnormal entries, such as blank or unreasonable values (e.g., the total floor area is 0m^2) are deleted. The properties with multiple records may have two different reasons, first, the records are purely identical data, so only one of these remains; second, the results are records assessed at different years that may reflect the changes in building conditions, so the records from the latest inspection date are used here.

The next step is to reorganise the classes of categorical data as inconsistencies exist when different inspectors create the records. The redundant categories may cause unnecessary waste on training time and computation power. Among all the records listed in table 1, the ones marked with an asterisk (*) are the records that require amending before using as inputs. The original lightning descriptions are texts in one format describing the proportion of low energy lighting. This study removed the texts, so the percentage of low energy lights is used as numeric data input. The rest of the classes follow similar re-categorising rules. In general, the EPCs contain classes that describe similar characteristics that can be combined (e.g., partial double glazing and some double glazing). There are also issues with formatting, language being used and units that need amending to be unified.

3.1.2. *Google street view images*. With the full addresses recorded in the EPC, it is possible to query the GSV images through a helper API. The API uses geo-coordinates or full addresses detailed to house numbers to extract the target GSV images with user-controlled photo specifications. An example of GSV obtained for the case study area is shown in the left image in Figure 1. To reduce the amount of irrelevant visual information, an object detection algorithm, YOLOv5, is applied to the extracted images.

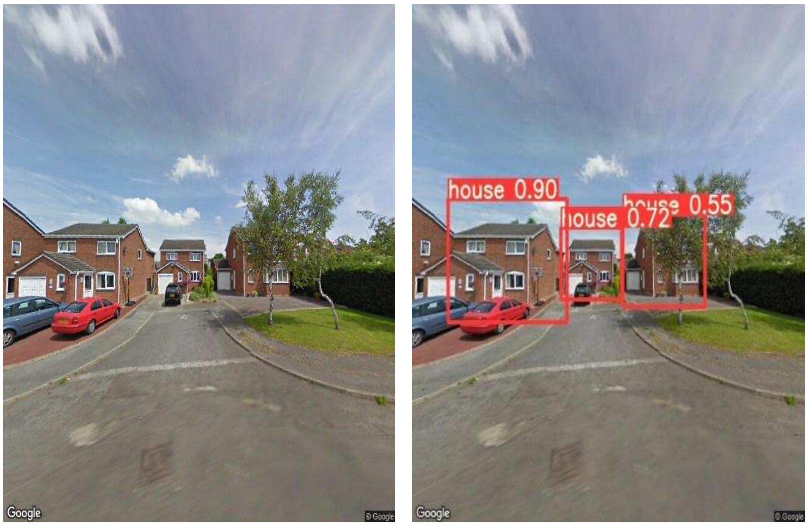

Figure 1. Example GSV image (left); Detection results for the example image (right)

A custom model is trained specifically for this study using over 800 manually labelled GSV images with bounding boxes. A weight is calculated to help determine the possibility of whether the detected pixels belong to a house feature. The example image input and its detection result are illustrated in Figure 1. The algorithm has successfully detected visual contents that look similar to houses. The value on the top of the boxes shows how likely the detected feature is a house. The evaluation results using average precision (AP) also suggest that the model is sufficient for this study's aim. AP is a commonly used metric in object detection when the detection task is set for a single class. It compares the ground-truth bounding box with the detected ones and produces a single value ranging from 0 to 1 [30]. The AP for the custom model for this study stays relatively constant after training for around 110 epochs and has resulted in an AP of around 0.8.

Once the houses in each GSV image are detected, these houses are saved as individual images as Figure 2. However, as the queried property should be a single building, it is necessary to select one of them as the target property when multi-detection appears. This step has taken the patterns found for houses' energy performance in the same neighbourhood, discussed in Section 2 [5], into consideration, so even if the properties next to the actual queried house are selected for prediction, the bias should be limited. These detected images in various sizes provide the possibility to determine the target house for the following energy performance prediction. If multiple houses are detected in the same image, the largest house detected is used for the following prediction.

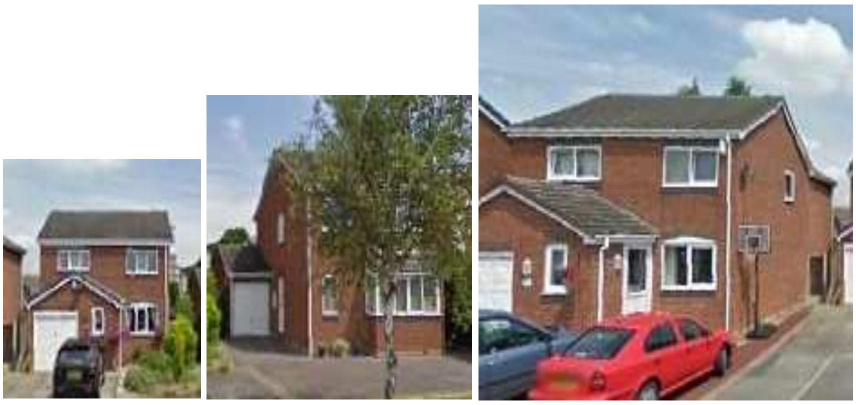

Figure 2. Example cropped images from multi-detection

Selecting the maximum cropped images have resulted in various sizes of input data. Paddings are added at the bottom to ensure all images have the same ratio of width and height (Figure 3), followed by a step to resize all the images to be 128 × 128 for the CNN model.

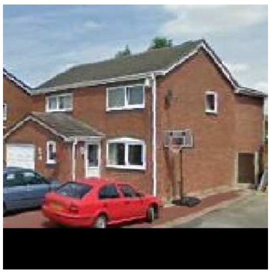

Figure 3. Example result after padding and resizing

3.2. Deep multimodal learning

The overall structure of the multimodal network is illustrated in Figure 4. The structure of MLP and CNN are selected for this study. Unlike applications using CNN and MLP models individually, the multimodal algorithm does not process each input completely through the entire algorithm to produce a result prediction, the process is designed to stop at the intermediate step, when both CNN and MLP produce four neurons from the inputs, so they can be concatenated and connected with the following fully connected layers for the final prediction.

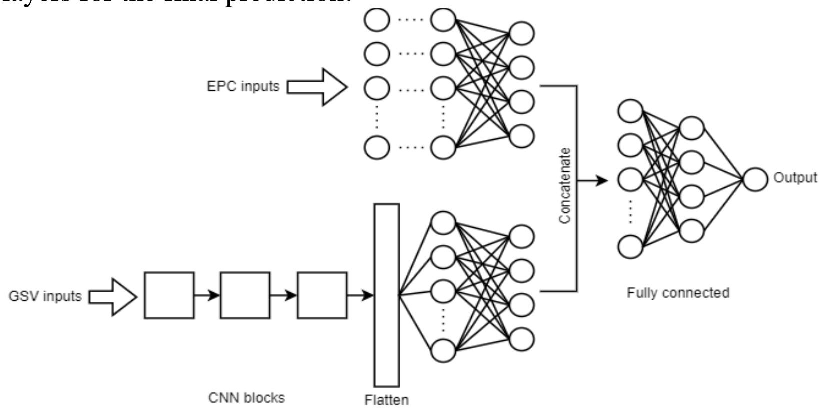

Figure 4. Illustration of the structure of the deep multimodal learning

The data extracted from the selected EPC records (Table 1) are used as input into an MLP model for training and predicting the building energy demands. A multi-layer perceptron is one of the most fundamental but powerful machine learning algorithms [31]. As the upper stream of Figure 4 illustrates, the MLP is a feedforward algorithm that calculates a function to connect each input with a neuron for calculating the outputs.

The CNN model is designed to process the visual modality. The CNN model is special because of its unique convolutional layer that processes the image input and produces representative patterns as result [32]. The input GSV images follow the lower stream of processing illustrated in Figure 4. They first go through three blocks of layers, each containing a convolutional layer, an activation layer (ReLu), and a max-pooling layer to create feature maps that reflect the most representative features inside of each scan. A flatten layer is added after the three CNN blocks to connect the model with fully connected dense layers and store the processed information in four neurons.

At this stage, both GSV inputs and EPC inputs are processed through their respective streams of learning and all result in a form of four neurons. A concatenate layer is added, so both the processed inputs become combined inputs containing features from both the CNN and EPC. Finally, dense layers are added to predict the residential energy consumption.

4. Case study: Barnsley, England

The case study was conducted with residential properties in the central region of Barnsley, UK. Overall, there are 11,740 EPC records found to be downloadable for central Barnsley. More than 15% of properties are found to be associated with multiple entries, filtering and reorganising the data is essential. By following the steps explained in the methodology section for data preparation, the input data for this case study contains 20,031 GSV images and their corresponding EPC records for 5,933 properties in central Barnsley. Property addresses that are unable to match with a GSV image are emitted for this study. Among these properties, 74.43% are houses (H), 16.3% are flats (F), 8.74% are bungalows (B), and the remaining 0.53% are maisonette (M). Most of the properties have an estimated energy consumption between 184 and 300 kWh/m^2 per year with a mean value of 224 kWh/m^2 per year. The average energy usage in central Barnsley is slightly lower than the national average. According to the official statistics, the mean energy uses for all existing properties in England in 2021 is around 267 kWh/m^2 per year [33]. There are also extreme cases where the recorded energy consumption is above 1,000 kWh/m^2 per year, but no clear associations are found with any recorded building characteristics.

Table 2 summarises the main type of building elements for each type of property. Similarities are found across different building types, especially in the choice of heating and fuels. This pattern is aligned with the status quo of existing properties in England, that according to the latest report by BEIS [1], gas is and has remained the predominant source of heating since around 1990. The main difference across building types in Central Barnsley can be seen in the elements that are related to the property structures: the floor, walls, and roofs.

Table 2. Most frequent building features in EPC for properties in central Barnsley.

Type	Floor	Window	Wall
B	Solid, no insulation;	double	Filled cavity
F	(Other dwellings below)	double	Insulated cavity
H	Suspended, no insulation	double	Filled cavity
M	(Other dwellings below)	double	Solid brick, no insulation

Type	Main heat	Roof	Fuel
B	Boiler and radiators	Pitched, 250 mm insulation	Gas
F	Boiler and radiators	(Other dwellings above)	Gas
H	Boiler and radiators	Pitched, 200 mm insulation	Gas
M	Boiler and radiators	Pitched, 100 mm insulation	Gas

5. Results and discussion

Three different predictions were conducted to compare and testify whether a deep multimodal network performed better in prediction accuracy compared with mono-modal or unimodal networks. More details will be discussed in the following. The three predictions are:

a. A mono-modal network only uses EPC records for residential energy prediction.
b. A mono-modal network only uses cropped GSV images for residential energy prediction.
c. A multimodal network uses both EPC records and cropped GSV images for residential energy prediction.

5.1. Network a: with EPC records

The MLP model has run for 100 epochs with a batch size of 32. The evaluation loss became relatively stable after around 95 epochs. The model reached a result of mean absolute difference of 15.45 kWh/m2 per annum between the predicted energy consumption and the value of current consumption recorded in the corresponding EPC, with a standard deviation (std) of 32.08.

5.2. Network b: with cropped GSV images

With 20,057 cropped and padded GSV images, the CNN model has been trained for 60 epochs with a batch size of 32. The designed CNN model processed the input GSV images and attempted to select the important features for prediction. Figure 5 is an example feature map created from the intermediate stage of the CNN model. Although no clear objects were detected, by comparing between the input layer and max-pooling layer, the model highlighted the outline of the building and the edges surrounding the building elements, i.e., roof, door, and window, for prediction. The final resulting mean absolute difference is 0.06 kWh/m^2 per annum between the predicted energy consumption and the value of current consumption recorded in the accordance EPC, and an std of 0.06.

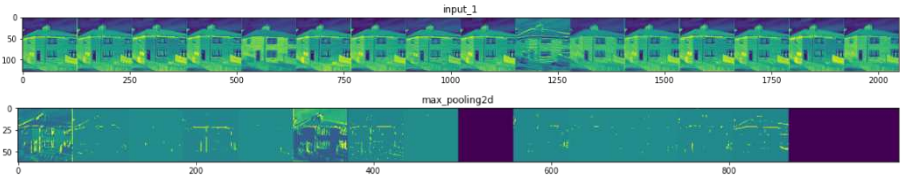

Figure 5. Example feature maps from the trained CNN model

5.3. Network c: multimodal network

The combined model has been trained for 100 epochs with a batch size of 32. The final result mean absolute difference between predicted and recorded values dropped to 0.01 kWh/m^2 per year.

6. Conclusion

This study aims at exploring the potential of using a multimodal deep learning network for residential energy consumption prediction. Benefiting from deep learning algorithms' capability in processing complex unstructured data, and the development of sensory systems, there is a growing trend of adopting deep learning that accepts multiple streams of data inputs. This study adopted the idea and developed an algorithm combining the structure of MLP and CNN algorithm to accept both text and visual inputs from EPCs and GSV images. By comparing with the unimodal networks, although all three networks can predict the potential residential energy consumption within a satisfying range of accuracy, the multimodal deep learning network further decreased the prediction error. This approach provides an accurate prediction of residential building energy consumption on a large scale. The results can provide insight into the energy performance of residential buildings without accessing to meter readings. A proper understanding of the existing housing stocks can help guide the necessary retrofit schemes and avoid unnecessary costs.

Using image data and trainable CNN networks creates feature maps to implicitly encode information on the structure of the building, as well as colour information and any urban furniture that may appear in the images. Through the training of the CNN component of the model, concatenated with the EPC-based model, insight into the interactions between these properties is learned. However, issues with available data, especially the inconsistencies in EPC data, still exist. Other data may be introduced as extra modalities for further studies to reduce the uncertainties. For example, thermal images may be helpful to use as a representation of the building insulation conditions. The thermal data may also be an indicator of where the thermal bridges exist. Lidar point cloud data could also be a potential modality. It may be used to reconstruct 3D representations of buildings to provide details, unlike the LOD1 GIS products, of the overall building structure and shape, and the allocations of elements including doors and windows. However, the spatial coverage of these potential modalities are all limited, which would be the main task to be overcome in future studies.

Acknowledgement

W Ward was supported by EPSRC Grant EP/V012053/1 and by Towards Turing 2.0 under the EPSRC Grant EP/W037211/1 and The Alan Turing Institute.

References

[1] BEIS. 2021. *National Statistics: Energy consumption in the UK 2021*. URL: https://www.gov.uk/government/statistics/energy-consumption-in-the-uk-2021

[2] Schofield B, Lewis C. COP26: Facing up to the challenge of retrofitting homes. *BBC News*. 2021 Nov 10; URL: https://www.bbc.co.uk/news/uk-england-northamptonshire-59227228

[3] Bourdeau M, Zhai X qiang, Nefzaoui E, Guo X, Chatellier P. 2019. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*. Elsevier Ltd **48**

[4] Wenninger S, Kaymakci C, Wiethe C. 2022. Explainable long-term building energy consumption prediction using QLattice. *Appl Energy* **308** 118300

[5] Rosser JF, Boyd DS, Long G, Zakhary S, Mao Y, Robinson D. 2019. Predicting residential building age from map data. *Comput Environ Urban Syst* **73** p 56–67.

[6] Department for Levelling Up Housing and Communities, Ministry of Housing Communities & Local Government. 2021. Guidance - Energy Performance of Buildings Certificates: notes and definitions. URL: https://www.gov.uk/guidance/energy-performance-of-buildings-certificates-notes-and-definitions

[7] Coyne B, Denny E.2021. Mind the Energy Performance Gap: testing the accuracy of building Energy Performance Certificates in Ireland. *Energy Effic* **14**(6)

[8] Zeppelzauer M, Despotovic M, Sakeena M, Koch D, Döller M. 2018. *Automatic Prediction of Building Age from Photographs*. In: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval. New York, NY, USA: ACM p 126–34

[9] Koch D, Despotovic M, Leiber S, Sakeena M, Döller M, Zeppelzauer M. 2019. Real Estate Image Analysis: A Literature Review. *J Real Estate Lit* **27**(2) p 271–300.

[10] Campbell A, Both A, Sun Q (Chayn). 2019. Detecting and mapping traffic signs from Google Street View images using deep learning and GIS. *Comput Environ Urban Syst* **77**.

[11] Hara K, Le V, Froehlich JE. 2013. *Combining crowdsourcing and Google Street View to identify street-level accessibility problems*. In: Conference on Human Factors in Computing Systems – Proceedings p 631–40.

[12] Hjortling C, Björk F, Berg M, Klintberg T af. 2017. Energy mapping of existing building stock in Sweden – Analysis of data from Energy Performance Certificates. *Energy Build* **153** p 341–55

[13] Summaira J, Li X, Shoib AM, Li S, Abdul J. 2021. Recent Advances and Trends in Multimodal Deep Learning: A Review *arXiv Prepr*

[14] Bayoudh K, Knani R, Hamdaoui F, Abdellatif M, Khaled B, Hamdaoui F, et al. 2021. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. *Visual Computer*

[15] Chen Y, Hong T, Piette MA. 2021. Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. *Appl Energy* **205** p 323–35.

[16] Mastrucci A, Baume O, Stazi F, Leopold U. 2014. Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam. *Energy Build* **75** p 358–67.

[17] Biljecki F, Sindram M. 2017. Estimating building age with 3D GIS. *Ann Photogramm Remote Sens Spat Inf Sci* **4** p17–24.

[18] Despotovic M, Koch D, Leiber S, Döller M, Sakeena M, Zeppelzauer M. 2019. Prediction and analysis of heating energy demand for detached houses by computer vision. *Energy Build* **193** p 29–35

[19] Huang Y, Chiang C-H, Hsu K-T. 2018. Combining the 3D model generated from point clouds and thermography to identify the defects presented on the facades of a building

[20] Lu X, Memari A. 2019. Application of infrared thermography for in-situ determination of building envelope thermal properties. *J Build Eng* **26**.

[21] Krüger A, Kolbe TH. 2012. Building analysis for urban energy planning using key indicators on virtual 3D city models - the energy atlas of Berlin. *ISPRS - Int Arch Photogramm Remote Sens Spat Inf Sci* **XXXIX-B2** p 145–50.

[22] Nouvel R, Mastrucci A, Leopold U, Baume O, Coors V, Eicker U. 2015. Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support. *Energy Build* **107** p 204–12.

[23] Zuhaib S, Schmatzberger S, Volt J, Toth Z, Kranzl L, Eugenio Noronha Maia I, et al. 2021. Next-generation energy performance certificates: End-user needs and expectations. *Energy Policy* **161**

[24] European Commission. Energy Performance Certificates.URL: https://ec.europa.eu/energy/eu-buildings-factsheets-topics-tree/energy-performance-certificates_en

[25] BRE. 2014. The Government' s Standard Assessment Procedure for Energy Rating of Dwellings. Garston. URL: http://www.bre.co.uk/filelibrary/SAP/2009/SAP-2009_9-90.pdf

[26] Crawley J, Biddulph P, Northrop PJ, Wingfield J, Oreszczyn T, Elwell C. 2019. Quantifying the measurement error on England and Wales EPC ratings. *Energies* **12**(18)

[27] Burman E, Mumovic D, Kimpian J. 2014. Towards measurement and verification of energy performance under the framework of the European directive for energy performance of buildings. *Energy* **77** p 153–63.

[28] Gebru T, Krause J, Wang Y, Chen D, Deng J, Aiden EL, et al. 2017. Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the United States. *Proc Natl Acad Sci USA* **114**(50) 13108–13.

[29] Yuan J, Cheriyadat AM. 2016. *Combining maps and street level images for building height and facade estimation*. In: Proceedings of the 2nd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics, UrbanGIS

[30] Redmon J, Divvala S, Girshick R, Farhadi A.2016. *You only look once: Unified, real-time object detection*. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. P 779–88

[31] Géron A. 2017. *Hands-on machine learning with scikit-learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Sebastopol

[32] Ian Goodfellow, Yoshua Bengio AC. 2016. *Deep Learning Book*. MIT Press.

[33] Ministry of Housing Communities & Local Government. 2021. Energy Performance of Buildings Certificates Statistical Release January to March 2021 England and Wales. *Gov.Uk*. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985584/EPB_Cert_Statistics_Release_-_Q1_2021.pdf