# Classifying City Level Housing Materials by Using Google Street View Images and Deep Learning

**Course Name：** Urban Analytics Group Project

**Group Members：** Quan Nie, Jingying Xu, Yinghu Chang, Changyu Han

**Advisors：** Professor Qunshan Zhao, Professor Fan Zhang

**Word Count:** 5533 (without reference)

**Date:** 2021.4.26.

## Abstract

The energy efficiency performance of urban housing properties is closely correlated to various regional environmental and societal issues, such as climate change and fuel poverty. Due to the multiple limitations on relevant building data set, measuring and recording housing materials in a certain district is relatively costly and hard to complete via traditional surveying methods like questionnaire or interview. The emerging analytical tools introduced from computing science and AI (Artificial Intelligence), however, are providing necessary theories as well as methodologies in understanding, identifying and measuring abundant elements of complicated urban built environment, making it more possible to handle such classification and quantification work. In this report, we will try to identify and then classify materials of housings' external wall in the City of London area through training and employing a deep learning model from the scratch. Enabled by new forms of big data (SVIs, street view images) and an advanced convolutional neural network — Mask R-CNN, we utilized street view images from the study area to train a model. Then we employed our model to analyze more housing materials of building images captured by Google Street View service in the study area, providing an overall pattern of housing material composition at the street/district level. Finally, the trained model is used to estimate the energy efficiency of housing materials in the test set area.

**Keywords:** Energy Efficiency, Google Street View, Housing Material, Mask R-CNN

# 1. Introduction

Housing energy performance as well as efficiency has long been one of the main focuses on the urban agenda (Sunikka-Blank et al., 2012) because of its potential but strong links with a number of key environmental and societal issues (Missoum et al., 2014), such as urban climate change and fuel poverty. In most prosperous post-urbanized districts, identifying and measuring the multi-dimensional concept of housing energy performance/efficiency (Cerin, Hassel & Semenova, 2014) is costly and unavailable in a context of a huge range of environmental elements and complicated building types. Therefore, many previous researches related to this theme mainly focused on housing's internal attributes or indicators, which normally have a direct relationship with the energy consumption amount or cost, such as the acreage of surveyed housing properties (see Cerin et al., 2014 as an example). In other words, limited by previously inaccessible or incomplete data set (Wang & Vermeulen, 2020) and expensive survey costs, some more extensive quantification projects, including identification, classification and measurement of built environment elements, were relatively hard to be employed in terms of urban and architectural studies (Batty, 2019) through traditional methodologies (Wang & Vermeulen, 2020).

In the past few years, the rapid rise and popularity of urban analytics has brought more methodologies that enable shifts from previous relatively "data-scarce" (Kitchin, 2013) studies to more accurate, quantified and complicated models (Wang and Vermeulen, 2020). "to measure the unmeasurable" (Batty, 2019) has also become a loud slogan,

symbolizes the much richer data sources and more automated and precise analytical tools that can be utilized in the field of urban studies. Combining emerging big data (e.g., street view images) with computational methods (e.g., machine learning algorithms), scholars now are able to pursue enhanced geographical and spatial understandings (Kitchin, 2013) toward our living environment (see LeCun et al., 2015 as an example) by implementing quantitative and replicable work. Therefore, in this report we will try to connect housing energy performance/efficiency with a previously "unmeasurable" factor, the housing materials (and their composition).

There are two main concerns that drive us to explore this topic of housing energy performance: climate change agenda and rising fuel poverty (Gonzalez-Perez & Leonard, 2017; Burlinson, Giulietti & Battisti, 2018). Generally speaking, various types and percentages of housing materials always have largely different level of thermal transmission rate (Pasichnyi et al., 2019), indicating different level of fuel consumption to maintain indoor temperature of these buildings, especially in the case of extreme weather (Missoum et al., 2014). For instance, skyscrapers with huge glass walls often have to consume much more energy to adjust its indoor temperature to a feasible degree in summer or winter, resulting in significant negative impacts on urban climate change like the emission of greenhouse gas and heat island effect (Gonzalez-Perez & Leonard, 2017). The other main concern correlated with housing materials and energy performance has been gradually recognized as "The elephant in the energy room" (Burlinson, Giulietti & Battisti, 2018), that the rising fuel poverty among European

countries may cause worse level of social equality and just. Such invisible and unintuitive factors of housing quality (i.e., the material-related energy performance) are likely to consume more unaffordable fuel costs out of deprived residents' control (Fabbri, 2015). In conclusion, these two concerns reveal the need to develop a model at satisfying accuracy to identify and calculate housing material composition of urban housing properties for further estimation of energy performance.

In following parts, this report will explain how we try to develop a model step by step to identify, classify and calculate the composition of materials of buildings' external wall in the study area (i.e., City of London): For the background part we are going to briefly review valuable literatures as well as online technical guidance that offered us essential theoretical and analytical tools to handle this task. Then there will be an introduction to our study area and how we collected, cleaned and filtered GSVs (i.e., Google Street View images) in order to get data ready before we constructed the instance segmentation model. Thirdly, the methodology part will present a detailed explanation about the whole procedure, involving analytical model training, building and application. After that, an overview of results and corresponding analyses will be visualized and offered. Finally, there will be a conclusion together with a discussion part assessing this model, evaluating the whole project and discussing more about future works.

## 2. Background

As a result of enhanced spatial understandings (Kitchin, 2013) enabled by deep learning algorithms toward our living environment (see LeCun et al., 2015 as an example), SVIs services such as GSVs make evolved visual recognition, classification and modelling accessible and replicable (Wang and Vermeulen, 2020).

These emerging studies are mostly presented in three ways:

(1) Primarily street view images have offered more opportunities to enable scholars to virtually access to their sites of interest (Curtis and Fagan, 2013) and record the changes of neighborhood elements (Hwang and Sampson, 2014), where in many cases official datasets are relatively inaccessible.

(2) Street view images can be further captured and identified via deep learning techniques and computer vision algorithms (LeCun et al., 2015) to recognize, extract and classify built environment factors, providing a rich data source to support assessing urban image elements like evaluating visual quality of streets (see Ye et al., 2019 as an example).

(3) Correlated with other forms of quantification or qualification method, built environment compositions extracted from SVIs can be used to do more in-depth correlation analyses on individual behaviours or outcomes in the neighbourhood such as crime studies (He et al., 2017).

According to the tasks required by our research theme, we mainly searched and read corresponding papers and technical documents for valuable and practical guidance to help employ adjustments and analytical configurations (i.e., the second and third mainstreams to apply convolutional neural network models).

For the academic literatures, as our current focus is on the data manipulation and model building, which is the second stream of using visual data to implement machine learning algorithms (see previous report and presentation for more details). Group members now narrowed our reading list down into researches that street view images are captured and identified via deep learning techniques and computer vision algorithms (LeCun et al., 2015). We were inspired by and learned from these scholars on how to recognize, extract and classify built environment factors, providing a rich data source to support assessing urban image elements and evaluating visual quality of a streets (see Ye et al., 2019 as an example).

As the technical guidance that helps build and refine our operating environment and configurations on a cloud service platform, we collected massive digital documents and manuals by producers of the Mask R-CNN and corresponding service providers. Furthermore, upon specific technical issues and problems we received generous and valuable guidance and help from our two advisors to get the testing model work.

# 3. Study area and data

## 3.1 Study area

The study area of this report is the City of London, which is located to the north of River Thames and covers an area of about 2.9 km$^2$, as shown in Figure 3-1. Today, this area gathers a large number of banks, stock exchanges and other financial institutions, with modern architecture style. Our Google Street View images were collected in this area.


Figure 3-1 Study Area

## 3.2 Data source and data cleaning

The study data came from Google Street View Maps. Get 58712 pictures with a resolution of 96 and a size of 400*400 pixels (width: 400, height: 400), 100 street scenes of timber house, and 2732 street view images with good camera angles.

**(1) Obtain street view images through road network data and Google Maps API**

We selected a block in London with 5 square kilometers area and used OSM (Open Street Map) to download this block road network data. After that, the projected

coordinate system of the road network data was changed through ArcGIS to make it consistent with Google Map to ensure the correct point coordinates. Through related analysis, exported the attribute table information of the picking point layer of the road network data (including longitude and latitude). Based on the processing of road network data, the map street view images were downloaded by calling the Google Map API (Application Programming Interface). Pictures were downloaded with viewing angles of 0 degrees, 90 degrees, 180 degrees, and 270 degrees at the same coordinate point.

**(2) Energy Performance Certificates (EPC) data acquisition**

Nearly 40% of energy consumption and carbon emissions in the UK come from the heating and use of our buildings. EPC refers to the certification issued by a professional energy assessor to a building. The data includes various indicators of the building, such as walls and roofs. It can make the energy efficiency of the building transparent and explain how to improve the energy efficiency of the building. Starting in 2007, except for a few exempted buildings, buildings must have EPC when they are constructed, sold, or leased. The EPC for domestic and non-domestic buildings is valid for 10 years, or until a newer EPC is produced for the building.

The EPC data for this study was taken from the Ministry of Housing, Communities & Local Government official website. Through the search, the EPC data set of the target block was obtained to 5537 sample data. Each sample contained multiple variables,

such as ADDRESS, POSTCODE, INSPECTION_DATE, WALLS_DDESCRIPTION, etc.

**(3) Data cleaning and filter**

Data cleaning and filter were mainly based on ArcGIS and Python. In the Google Street View (GSV) image dataset, there were four photos with different angles at the same coordinate point, which meant that four times the photos needed to be processed, which brought us difficulties in the data processing. Therefore, using Python calculations, we reduced four photos of the same coordinate to one record. To link the street view pictures with the surrounding EPC points, we used ArcGIS to associate the EPC buildings within 20 meters of the street view camera position. EPC data points beyond a radius of more than 20 meters had been deleted. Finally, we obtained about 3975 data samples merged with EPC and street view cameras.

### 3.3 Manual labeling and sample supplement

Based on the WALLS_DESCRIPTION variable in the EPC data set, we found that there are more than 8 types of housing materials, namely System built, Average thermal transmittance X (X was a different number), Granite or whin, Sandstone or limestone, Sandstone, Timber frame, Cavity wall, Solid brick. Housing materials could only be labeled by the subjective vision and EPC data, so we had reclassified the categories. System built and Average thermal transmittance X (X was a different number) were classified as glass category. Granite or whin, Sandstone or limestone, and Sandstone

were considered as stone categories. Timber frame was used as timber. Cavity wall and Solid brick were used as brick categories. Based on the calculation requirements of the subsequent model, in addition to the above materials, other materials or substances on the picture were classified as others.

Next was the manual labeling. The tool we used was VGG Image Annotator (VIA), which was a very convenient offline labeling software. Through VIA, we selected 400 pictures from 3975 samples for manual labeling. However, after finishing the labeling, we found that in the five categories, the samples of timber material were particularly small. The main reason might be the lack of buildings of this material in the block we selected. To make up for this shortcoming, we downloaded 100 photos of timber buildings in other blocks of London and hand-labeled them.

Finally, our training data set consists of 500 pictures of handmade house material labels. The house material labels were classified into five categories, namely glass, stone, timber, brick, and others. In addition, by adjusting the angle of the better Google Street View camera, 2372 clear street view pictures of the building were obtained as a verification set, and energy efficiency was estimated based on this.

# 4. Methodology

## 4.1 VGG Image Annotator

This project uses deep learning methods to identify and classify housing materials and

the first step is to label the images dataset manually. We chose to use VGG Image Annotator (VIA) to annotate housing materials, which is an open-source image annotation tool. This tool does not require installation, can be used online and offline. It is also easy to learn and operate, and can mark rectangles, circles, ellipses, polygons, points and lines. We used polygon annotation tool to mark in images and set house materials into four categories, namely glass, timber, brick and stone, as shown in the Figure 4-1. After labeling, it can export the annotation as json file format which facilitate to the subsequent model.

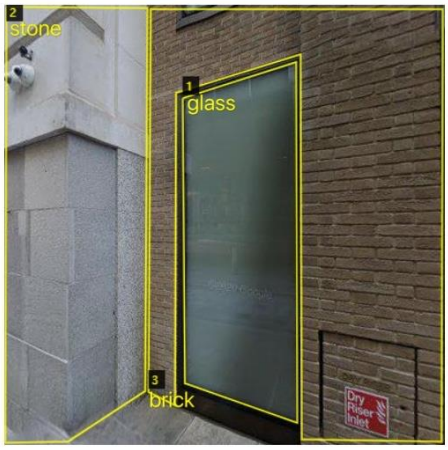


Figure 4-1 Label

### 4.2 Mask R-CNN

There are many deep learning algorithms that can achieve instance segmentation. In this project, we chose Mask R-CNN to train on our housing material dataset, because this algorithm has the characteristics of high accuracy, fast speed, simple network structure and flexible application. Mask R-CNN is a convolutional network based on

the Faster R-CNN architecture which achieves high-quality instance segmentation while performing target detection. It adds a prediction mask branch on the basis of Faster R-CNN to predict the object in parallel and replaces the Region of Interest (ROI) Pooling module with a more accurate ROI align module. Therefore, this algorithm can effectively detect the objects in the image and generate a high-quality segmentation mask for each instance in the meantime. Compared with Faster R-CNN, Mask R-CNN is easier to implement and train, and the speed is also faster, which is 5 fps (Ahmed et al., 2020,). The framework of Mask R-CNN contains two stages which both connected to the backbone structure. The first stage is scanning the input image and generating proposals for areas which may contain an object, the second stage is classifying the proposals according to the first stage and generating bounding boxes and masks according to the classification (Ahmed et al., 2020).

Mask R-CNN consists of these modules: Backbone, RPN and RoI-Align (as shown in Figure 4-2). The backbone is a standard convolutional neural network, and Mask R-CNN use the backbone model like ResNet 101 or ResNet-FPN architecture to extract features from input images. Then the Region Proposal Network (RPN) which is a light weight neural network predict if it contains the object. The area scanned by the Feature Pyramid Network (FPN) with a sliding window is called an anchor, and each anchor will generate two outputs: the first one is the anchor category is foreground or background, where the foreground one is called positive anchor; the second one is the evaluation of parameters of positive anchors boxes, which is used to fine-tune the boxes.

In this process, FPN will generate the feature maps which called Region of Interest (RoI). RoI-pooling is used to down-sample the important features from feature map in Faster R-CNN. To avoid the problem of misalignment between the extracted features and the region of interest caused by the RoI-pooling operation, RoI-align is introduced to output multiple bounding boxes and convert them to a fixed dimension. In this step, anchors boxes are used to detect multiple specific objects in an image and capture their scale and aspect ratio. Since Mask R-CNN can generate a large number of anchor boxes, Intersection over Union (IoU) is used to remove unnecessary boxes.

IoU = Area of the intersection / Area of the union

That is to say, IoU measures the degree of overlap between the real bounding box (ground truth) of the object to be detected in the image and its prediction box. The value of IoU is between 0 and 1, the larger the value, the higher the degree of overlap. A value of 0.5 means there is a 70% overlap between the two regions. Generally speaking, when IoU is greater than or equal to 0.5, the area is considered as an area of interest. When IoU equals to 1, the predicted bounding box completely overlaps the real one. Non-Max Suppression will keep the bounding box with the highest IoU value.

After that, there will be a branch in this module to generate a mask for each area that contains the specific object. This step is carried out through Fully Convolution Network (FCN), and it achieves semantic segmentation rather than instance segmentation. This is because each RoI value corresponds to one object, it is only necessary to perform semantic segmentation on it, which is equivalent to instance segmentation. Training and prediction of the model are different in this step. During prediction, the result of the

classifier needs to be obtained first and then passed to the mask branch to get the mask, whereas the classifier and the mask are performed at the same time during training.

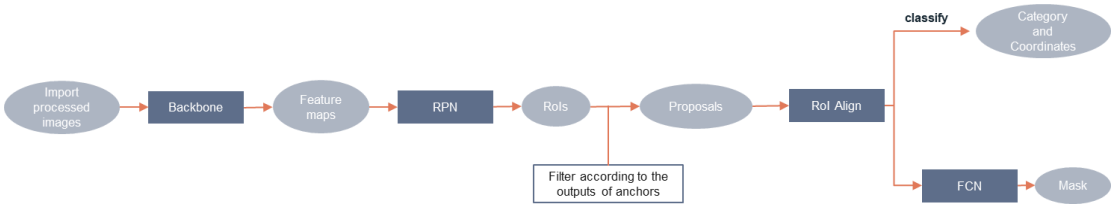


Figure 4-2 Main structure of the model

## 4.3 Evaluation method

The performance of the trained model can be evaluated from many aspects, including running time and memory usage, but the most important evaluation still depends on the evaluation indicators. Common evaluation indicators for instance segmentation include accuracy, precision, recall, etc. Although different indicators have different calculation methods, they are essentially calculated based on the confusion matrix. Confusion matrix is essentially a table that shows the classification results of a classification model, as shown in Table 4-1.

*Table 4-1 Confusion matrix*

| Confusion matrix | | Prediction | |
|---|---|---|---|
| | | Positives | Negatives |
| True | Positives | TP | FN |
| | Negatives | FP | TN |

Among them, TP (True Positive) represents the true value is positive, and the model

considers the number to be positive; FN (False Negative) indicates that the true value is positive, and the model considers the number to be negative; FP (False Positive) indicates that the true value is negative, and the model considers the number to be positive; TN (True Negative) means the true value is negative, and the model considers the number to be negative. TP and FP are obtained through IoU, and the IoU is calculated to determine whether a detection result is true or false. The commonly used threshold is 0.5, which means that if IoU is greater than 0.5, it is a true positive.

Accuracy is the ratio of the number of correctly classified samples to the total number of samples. The expression in confusion matrix is as follow:

$$\text{Accuracy}= \frac{TP+TN}{TP+TN+FP+FN}$$

It can be used to measure the ability of the classifier to detect the entire sample, but when the sample has multiple categories and the distribution of each category is extremely unbalanced, there will be a situation where the accuracy is very high but the model detection performance is very poor. As a result, it is usually used to evaluate the overall accuracy of the model, and cannot fully evaluate the performance of the model. In addition, in the Mask R-CNN algorithm, accuracy is mainly to determine whether the object is detected while precision is to determine whether the mask precisely covers the object. Hence, precision is mainly used to evaluate the model performance.

Precision is the proportion of positive samples that are correctly predicted to all samples that are predicted to be positive. The expression is as follow:

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall is the proportion of positive samples that are correctly predicted to all samples that are actually positive, which focuses on how many true positives are correctly detected. The expression is:

$$\text{Recall} = \frac{TP}{TP+FN}$$

Precision focuses on whether the identified positives are reliable while recall pays attention on whether the model is sensitive to positives. The higher the two indicators, the better the performance of the model. However, they will affect each other, that is, when the precision is high, the recall is low. Therefore, the two indicators need to be combined to evaluate model performance and the Precision-Recall curve is obtained. It takes Precision as the vertical axis and Recall as the horizontal axis. In the P-R Curve, the closer to the upper right corner, the better the model performance. Since IoU and the confidence threshold will affect Precision and Recall, which will cause the curve to change, it is necessary to limit the value of IoU and the confidence threshold. For IoU, its value can be fixed to 0.5, but the confidence threshold varies greatly among different models. Therefore, an evaluation parameter that can be applied to all models is required, that is, Average Precision. AP is the area under the Precision-Recall Curve, which is used to measure the quality of the model in each specific category.

In this project, because there was more than one category to be detected, the performance of each category should be considered comprehensively. Therefore, Mean Average Precision(mAP) was introduced, which was the average value of Aps in

multiple categories. This indicator was one of the most important indicators in the target detection algorithm and was suitable for algorithms that simultaneously detect object types and positions. Therefore, we used this indicator to evaluate the performance of our model. The value of mAP is between 0 and 1, and the larger the value, the better the performance of the model.

# 5. Results and analysis

## 5.1 Mask R-CNN implementation on the housing material dataset.

### (1) Inspect dataset and pre-processing

We divided the data set into training set, validation set, and test set according to the ratio of about 8:1:1. The number of images and labels was shown in Table 5-1.

*Table 5-1 Dataset statistics*

| | **Number of images** | **Number of labels** | | | |
|---|---|---|---|---|---|
| | | glass | stone | brick | timber |
| **Training set** | 386(78%) | 1735 | 453 | 465 | 75 |
| **Validation set** | 53(11%) | 305 | 218 | 37 | 2 |
| **Test set** | 55(11%) | 276 | 125 | 34 | 6 |

Before training the model, we preprocessed the images: load masks and category name generated by images and its JSON file, then computed the bounding boxes from masks. Due to the limited data set , we used two data augmentation methods to expand the

dataset. Firstly, flip images right/ left half the time. Secondly, add a Gaussian blur with a random sigma in the range 0 to 5.

**(2) Model training**

We used the VMware Engine training model on the Google Cloud Platform. The configuration was shown in Table 5-2.

*Table 5-2 Google VMware Engine settings*

| **Environment** | **Python 3 (with Intel® MKL and CUDA 11.0)** |
|---|---|
| **GPU** | NVIDIA Tesla K80 x 1 |
| **Machine** | 4 vCPUs, 15 GB RAM |

To reduce training requirements, we used transfer learning, that was, starting from the weight file trained on the COCO dataset. At this time, trained weights had learned a lot of natural image features. The Mask R-CNN model training process had two stages. Firstly, we froze the backbone layers and trained the heads (i.e., the layers that we did not use pre-trained weights from MS COCO) for 2000 iterations (20 epochs x 100steps), starting from a learning rate of 0.001. Secondly, we fine-tuned all layers by iterate 2000 times with learning rate equals to 0.0001. The training took less than 2 hours. The result of general loss of training and validation after the last iteration were below 2. (The general loss metric was the sum of the five losses—Region Proposal Network anchor classifier loss and bounding box loss, Mask R-CNN classifier loss, bounding box loss and mask binary cross-entropy loss (He et al., 2017).

**(3) Evaluation and Detection**

Instance segmentation performance on this task was measured by the VOC-Style mask AP50 (averaged over IoU thresholds of 0.5). It should be noted that the IoU of the mask and not the bounding box was calculated here. We calculated mAP for all type detections of each picture in the test set. Then took the average of all test set images to get the final mAP. The model had a val/test mAP of 0.44/0.38. Examples of the results of using this model to detect housing materials were shown in Figure 5-1.


Figure 5-1 Result of Mask R-CNN on house material dataset

It could be found that the best detection result is glass, and the model could accurately identify the glass part of the wall of other building materials. The recognition result of brick and stone material was shown as a complete mask in some test images. However, in the process of dataset annotation, due to irregular shapes of brick, stone, and timber

walls in some pictures (with glasses in them), the annotations would be split. Thus, in terms of practicality, that was, to identify the type of material, the identification result was accurate, but this would affect the evaluation of the model. After using the model to identify the building materials in the street view images, we counted the number of instances detected in each category and calculated the area percentage of different housing material categories in the entire picture by summing the mask pixels. The statistical results were used to analyze the distribution of housing materials in the study area and analyze energy performance.

### 5.2 The energy performance of housing material

#### (1) Housing material label distribution

By using the trained model, in the City of London, we supplemented the 2372 Google Street Images with better angles to test. We found it identified 18,947 housing material labels. As shown in Figure 5-2, the Glass tag occupies most of it, even about twice the sum of the other three. The number of Brick and Stone tags was similar (brick: 401; stone: 459), while the number of Timber tags was the least (timber: 101). As shown in Figure 5-3, among the 18,947 tags identified by the model, the area of them was counted. We find that the area of glass material accounted for 72% of the total area, brick material accounted for 10%, stone material accounted for 15%, and timber material accounted for 3%.

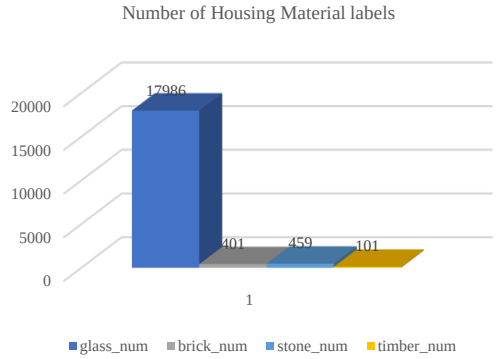


Figure 5-2 Number of labels for 4 types of housing materials

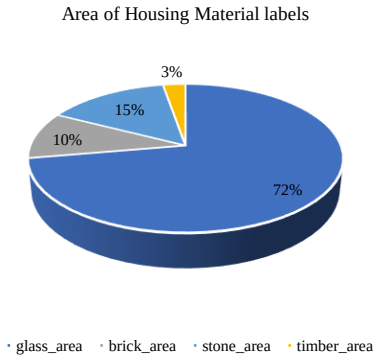


Figure 5-3 The area ratio of the four housing materials

As shown in Figure 5-4 to 5-7, We made statistics on the proportion of the material area on each picture, and found that the glass material in the street view picture mainly occupies the range of 0 to 20%. Brick material accounted for 0 to 8% of the total area of street view image recognition. Stones were distributed from 0 to 7%. Timber materials were concentrated in the range of 0 to 14%. It could be seen from this result that in the recognition process, there were many non-building objects in the street view image data set, such as cars, trees, streets, and sky.

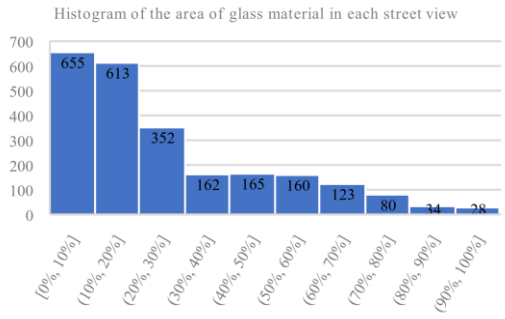


Figure 5-4 Histogram of the area of glass material in each street view

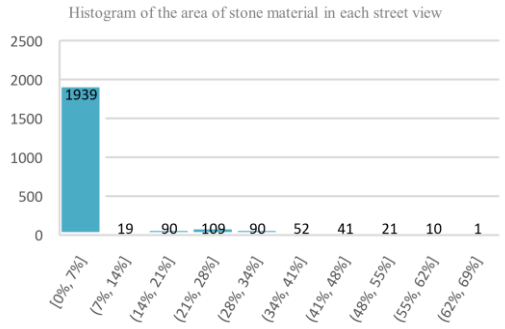


Figure 5-5 Histogram of the area of stone material in each street view

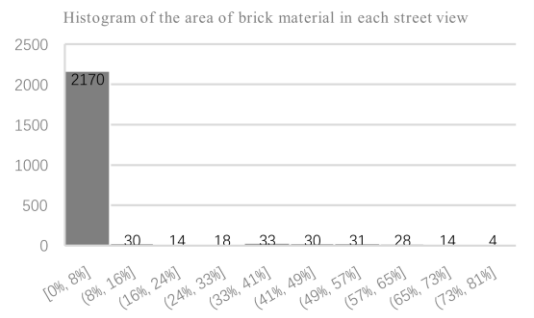


Figure 5-6 Histogram of the area of brick material in each street view

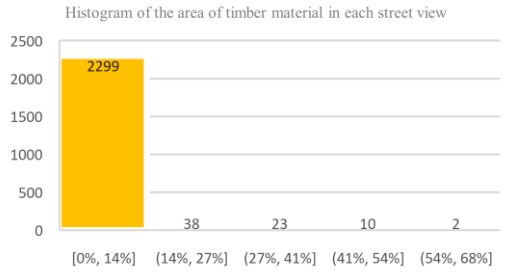


Figure 5-7 Histogram of the area of glass material in each street view

## (2) The energy efficiency of housing material

In general, our test set in the City of London area was at a medium level of Energy Efficiency. EPC had seven levels, from A (highest energy efficiency) to G (lowest energy efficiency). As shown in Figure 5-8 and 5-9, the rating content mainly included two aspects: firstly, it was the amount of energy used per square meter (lighting, heating, and hot water); Secondly, it was the level of carbon dioxide emissions. Since our study area was the City of London, the houses construction period of the building was not long. Ignoring the difference between each building of the same material, we calculated the energy consumption of house materials in the entire study area in an estimated way.

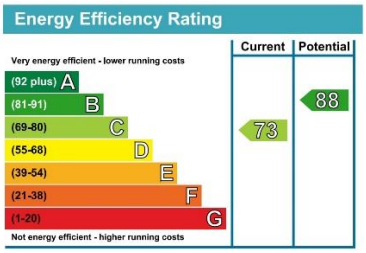


Figure 5-8 The example of Energy Efficiency Rating

*Picture Source:* https://justyy.com/archives/6885

| Element | Description | Energy Efficiency |
|---|---|---|
| Walls | Solid brick, as built, no insulation (assumed) | ★☆☆☆☆ |
| Roof | Pitched, no insulation (assumed) | ★☆☆☆☆ |
| Floor | Suspended, no insulation (assumed) | — |
| | Solid, no insulation (assumed) | — |
| Windows | Mostly double glazing | ★★★☆☆ |
| Main heating | Boiler and radiators, mains gas | ★★★★☆ |
| Main heating controls | Programmer, TRVs and bypass | ★★★☆☆ |
| Secondary heating | None | — |
| Hot water | From main system | ★★★★☆ |
| Lighting | Low energy lighting in all fixed outlets | ★★★★★ |

Figure 5-9 Energy Efficiency scoring standards

*Picture Source：* *https://find-energy-certificate.digital.communities.gov.uk/*

Since the Energy Efficiency rating was a complicated process, the house wall could not be used to judge the rating alone. Therefore, the Energy Efficiency rating results of the entire EPC data set of the City of London were used to estimate the Energy Efficiency of the house materials we had identified. As shown in Figure 5-10, we had obtained statistics on the house grades in the EPC files of the entire block, and there were almost no Energy Efficiency grades of A. In the house with glass material of the wall, most of the grades were C, followed by D and B. In houses with the brick material wall, most of the grades were C, followed by D and E. In stone houses, most of the grades were C, followed by D and B.

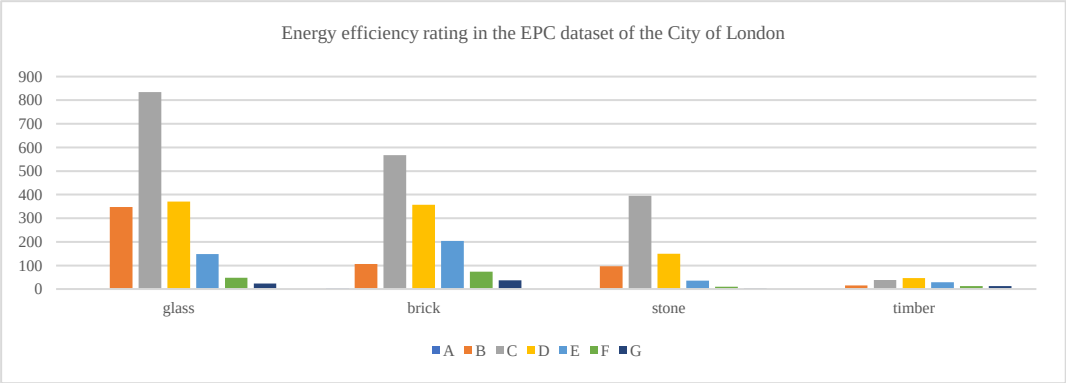


Figure 5-10 Energy efficiency rating in the EPC dataset of the City of London

As shown in Figure 5-11 to 5-13, the Energy Efficiency of the data set we identified was obtained after weighted analysis. From an overall point of view, in the test set, the Energy Efficiency grade was mainly C, followed by D and B. The glass material almost played a decisive role due to its powerful quantity. From the perspective of the 4 types of housing materials, except for timber, the remaining Energy Efficiency grades were mainly distributed in C. In general, our test set in the city of London area was at a medium level of Energy Efficiency.

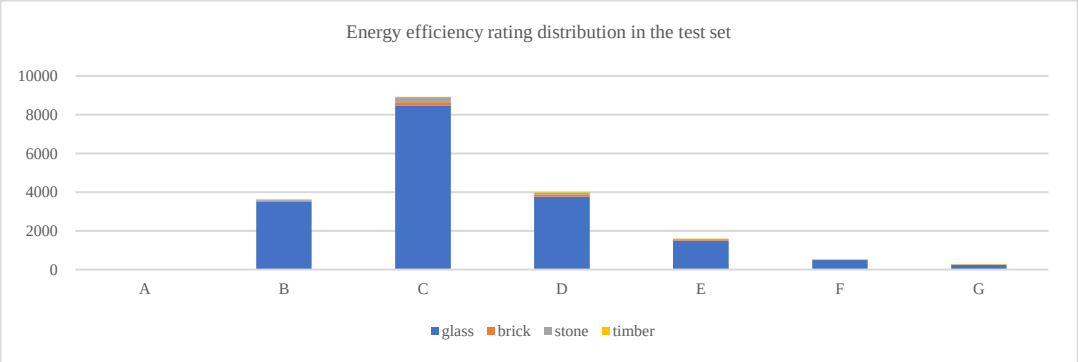


Figure 5-11 Energy efficiency rating distribution

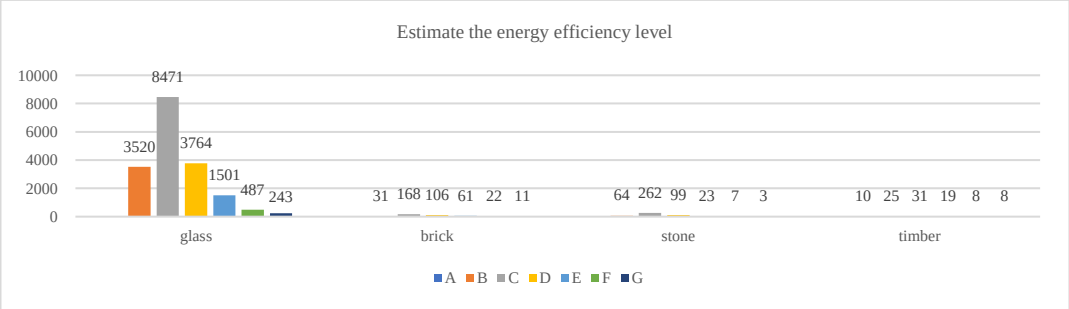


Figure 5-12 Estimate the energy efficiency level of 4 materials

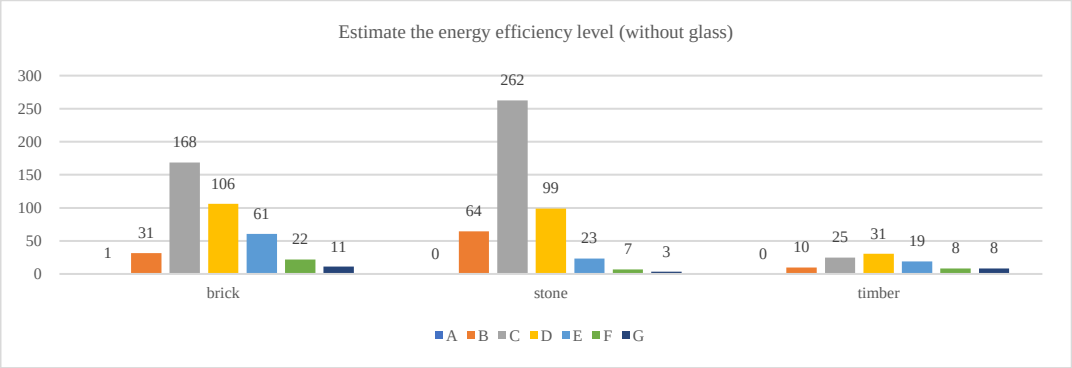


Figure 5-13 Estimate the energy efficiency level of 3 materials

## 6. Conclusion

This research aims to establish an instance segmentation model to identify the house materials in the city of London area and then analyze the energy performance of this area in combination with the EPC energy consumption standard. We used the Google Street View images of this area as the data source and labeled the house materials in the image, with the reference of the EPC data near the image's coordinates. The Mask R-

CNN model is built on this dataset, realizing the instance segmentation of house materials in street view images. And the model is used to detect the types of housing materials in well-angled street-view pictures of more than 2,000 locations in the city of London area and then combine with the EPC dataset to evaluate the Energy Efficiency of the house materials in this area. The study results show that the most detected glass materials in this area, followed by bricks and stones, and wood the least (unified with the labeled data). And the Energy Efficiency grade is mainly C, followed by D and B. In total, the study area is at a medium level of Energy Efficiency.

## 7. Discussion

There are still many limitations in the project, which need to be improved in the three points in future work. (1) In terms of data sources. The research area of this report is close to the CBD area of London. The style is a mainly modern architectural style reflected in the number of building materials is stone/glass > brick >> timber. In addition, Google Street View images are mainly taken on arterial traffic roads, and there is less camera point for l feeder highway and branch roads. Therefore, it is not easy to obtain a diverse sample of building materials in an area of about 2.9 square kilometers. In the future, those problems can be solved by expanding the scope of research. This will also provide conditions for analyzing the distribution of energy efficiency in units of different levels. (2) In terms of data labeling, brick, stone, and wood walls cannot be used as a complete label due to a large amount of glass embedded, which will impact model recognition and evaluation. How to label more accurately still needs to be

discussed. (3) In terms of model training, the current detection results on the glass category are more accurate than other types. The reason may be that other types of data have fewer labels than glass, and the existing sample diversity is too high, resulting in insufficient model learning. It can be solved by increasing the sample size in the future. Besides, introducing uncertainty into energy performance estimates and proposing methods to reduce uncertainty also need to be studied.

# References

Ahmed, B., Gulliver, T.A. & alZahir, S. 2020, "Image splicing detection using mask-RCNN", Signal, image and video processing, vol. 14, no. 5, pp. 1035-1042.

Batty, M. 2019, "Urban analytics defined", Environment and planning. B, Urban analytics and city science, vol. 46, no. 3, pp. 403-405.

Burlinson, A., Giulietti, M. & Battisti, G. 2018, "The elephant in the energy room: Establishing the nexus between housing poverty and fuel poverty", Energy economics, vol. 72, pp. 135-144.

Curtis, A. & Fagan, W.F. 2013, "Capturing Damage Assessment with a Spatial Video: An Example of a Building and Street-Scale Analysis of Tornado-Related Mortality in Joplin, Missouri, 2011", Annals of the Association of American Geographers, vol. 103, no. 6, pp. 1522-1538.

Cerin, P., Hassel, L.G. & Semenova, N. 2014, "Energy Performance and Housing Prices", Sustainable development (Bradford, West Yorkshire, England), vol. 22, no. 6, pp. 404-419.

Fabbri, K. 2015, "Building and fuel poverty, an index to measure fuel poverty: An Italian case study", Energy (Oxford), vol. 89, pp. 244-258.

Gonzalez-Perez, M.A. & Leonard, L. 2017, Climate change and the 2030 corporate agenda for sustainable development, Emerald, [Bingley], England.

Hwang, J. & Sampson, R.J. 2014, "Divergent Pathways of Gentrification: Racial Inequality and the Social Order of Renewal in Chicago Neighborhoods", American sociological review, vol. 79, no. 4, pp. 726-751.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. 2017. "Mask r-cnn". In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

He, L., Páez, A. & Liu, D. 2017, "Built environment and violent crime: An environmental audit approach using Google Street View", Computers, environment and urban systems, vol. 66, pp. 83-95.

Kitchin, R. 2013, "Big data and human geography: Opportunities, challenges and risks", Dialogues in human geography, vol. 3, no. 3, pp. 262-267.

LeCun, Y., Bengio, Y. & Hinton, G. 2015, "Deep learning", Nature (London), vol. 521, no. 7553, pp. 436-444.

Missoum, M., Hamidat, A., Loukarfi, L. & Abdeladim, K. 2014, "Impact of rural housing energy performance improvement on the energy balance in the North-West of Algeria", Energy and buildings, vol. 85, pp. 374-388.

Pasichnyi, O., Wallin, J., Levihn, F., Shahrokni, H. & Kordas, O. 2019, "Energy performance certificates — New opportunities for data-enabled urban energy policy instruments?", Energy policy, vol. 127, pp. 486-499.

Sunikka-Blank, M., Chen, J., Britnell, J. & Dantsiou, D. 2012, "Improving Energy Efficiency of Social Housing Areas: A Case Study of a Retrofit Achieving an "A" Energy Performance Rating in the UK", European planning studies, vol. 20, no. 1, pp. 131-145.

Wang, M. & Vermeulen, F. 2020, "Life between buildings from a street view image: What do big data analytics reveal about neighbourhood organisational vitality?", Urban studies (Edinburgh, Scotland), , pp. 1-22.

Ye, Y., Zeng, W., Shen, Q., Zhang, X. & Lu, Y. 2019, "The visual quality of streets: A human-centred continuous measurement based on machine learning algorithms and street view images", Environment and planning. B, Urban analytics and city science, vol. 46, no. 8, pp. 1439-1457.