XXVIII International Seminar on Urban Form
ISUF2021: URBAN FORM AND THE SUSTAINABLE AND PROSPEROUS CITIES
29th June – 3rd July 2021, Glasgow

Morphological characterization of landscape using context-rich geometrical features extracted from path centre lines

Karimi Kayvan[1], Eduardo Rico[2], Iacopo Neri[3]

[1] Space Syntax Laboratory, Barlett School of Architecture, UCL, UK

[2] Space Syntax Laboratory, Barlett School of Architecture, UCL, UK

[3] Institute for Advanced Architecture of Catalonia – IAAC

Abstract

Understanding the morphology of connective networks can reveal hidden aspects of our cities when correlated to geographical data. A strong body of work exists on examining the relationship between space syntax analysis and socioeconomic indicators, in some cases complemented with information on urban blocks and typologies. Other authors study street morphology either as the basis of generative algorithms or detecting global groups of urban patterns across many cities. However, other aspects of urban and landscape design, such as style or character, while widely theorised and used within generative algorithms, have not been subjected to systematic geospatial exploration. This study presents an analytical method that can adequately capture the geometrical nature of urban landscape networks and correlate it with its historical style or character. In our experiment, we extracted geometric features from OSM cleaned path centre lines for 40 parks in London. This was followed by a historical survey carried out for each park, labelling lines into six distinct historical-stylistic categories. Eight classifiers were trained and evaluated through a machine learning process predicting the historical categories of lines from their geometrical features. This was repeated in three consecutive tests with a growing degree of contextual features derived from neighbouring lines. For the last test, a bespoke technique including spatial and non-spatial clustering was used to identify morphologically coherent pieces of fabric which are fed into the classifier. Classifiers accuracy range from 55% in the first test to above 90% using context-aware geometric features. This suggests that the historical development of the urban or landscape fabric leaves a footprint in the map that can be unravelled by analysing its morphology. In future studies, these methods have a great potential to be expanded further towards investigating these findings in relationship with other more sophisticated, historically-dependent data and be applied to different types of networks.

Introduction

Urban Morphology (UM) and the understanding of cities departing from the study of their form has become a robustly grounded discipline with applications in multiple fields of research and practice and has recently gained traction due to the current developments in AI technology. (Dibble et al., 2019), borrowing from theories of biometrics, use the term Urban Morphometrics as a general method for the study of cities based on the analysis of elements such as streets, blocks and buildings, typically correlating this data with diverse socioeconomic and cultural aspects. This work has also found its way to software packages that generate urban patterns in industries such as urban design, computer games or cinema.

However, the field of landscape design has been left relatively unexplored within the field of urban morphology, particularly when it comes to the characterisation of its networks and their relation to the historical development of cities. Many patterns that nowadays define our urban environments such as star-shaped axis and curvaceous layouts were initially developed within garden and landscape design and were later adapted to the necessities of cities. Being able to study these patterns and how they have been translated to maps is an important research field for the field of UM which is still requiring further work. Our research addresses this gap, where we try to identify the relation between the morphology of landscape networks and their historical design character. We do so by extracting geometrical measurements from maps of several parks in London and studying how this data correlates with a categorisation of their historical style.

Background

(Oliveira, 2019) identifies the work of Conzen and Whitehand as setting the foundations of an analytical era of UM which becomes later cemented with the works of authors such as (Hillier and Hanson, 1984) and Batty and the development of computational geospatial methods. This work typically begins by extracting measured features from elements on the urban fabric which are then fed to algorithms that group them, sort them or correlate them to other sources of data. (Dibble et al., 2019) review existing literature and define a set of 9 main indicators (down from an initial set of 207) which enable them to classify 90% of urban settings by historical origin. (Netto et al., 2020) analyse block footprints in the shape of Nolli maps of downtown areas to find out similarities across cities. Working at a country scale, (Boeing, 2018) use UM methods to perform aggregate analysis of large pieces of road networks to obtain classifications of cities across the US. More recently (Fleischmann, 2019) has published a bespoke programming library focusing on UM.

While the previous techniques focus on the explicit analysis of vector data of shapes and network, current techniques of image processing have allowed the incorporation of other sources of information such as satellite or other raster-based datasets. (Moosavi, 2017) use convAE trained on Open Street Map (OSM) tiles of urban fabric across the world to extract global patterns of urbanization. (Golan Levin, 2017) scan the entire planet allowing people to discover land-use patterns similar to a given sample with their Terrapatern web app, opening up applications for activism and research. Other approaches relate geographical analysis with street view images or crowdsourced data. (Gonzalez and Alhasoun, 2019) use this approach to predict street typologies, (Law et al., 2019b) do similar work with house prices. (Nice et al., 2020) combine this approach with satellite view and rasterized maps to find relations between cities and (Scepanovic et al., 2021) use satellite images to predict urban vitality on a quarter-by-quarter basis. In some cases, these approaches are complemented with collaborative mapping and crowdsourcing data (Bubalo et al., 2019), (Streetscore, 2018) (Law et al., 2018) or (Redi et al., 2018).

Another group of authors would focus on the discovery of zones or segments of the urban fabric that present a certain similarity or coherence. (Schirmer and Axhausen, 2015) combine morphology and configurational approach to generate coherent clusters of urban blocks at the city scale. The authors manage to discover large regions of coherent character across the city, in this case, strongly governed by the syntax “contour” distribution. Based solely on syntactic data, (Law et al., 2019a) use community detection methods on a graph basis to identify coherent areas within the city based on different syntax measures. The authors identify which of these measures produce partitions that best match borough definitions in London and Amsterdam

However, these approaches don’t focus primarily on aspects of landscape networks nor address their historical and stylistic nature. Moreover, landscapes and parks are likely to have additions or details “nested” within a larger matrix of a different character, in some case composed of predefined syntax such as “axis + rotonda”. Current work in UM does not focus on this granularity of the landscape phenomenon. Our work tries to address these shortages and we propose a method that brings measurements specific for landscape design and also scan for elements or nested components hidden inside large parks.

Methodology

To develop this experiment, we use single-line drawings extracted from OSM (also called road centre-line drawings) selected from the 50 largest parks in London and we carry out a series of tests using geometric information extracted from the lines to predict the historical classification of the lines or entire sections of the parks (what we will call parklets). In each of these tests, we increase the amount of information that we draw from the context for each line with the expectation that richer context data will produce better prediction results. The underlying hypothesis is that methods that capture enough information on the geometries of these lines shall be able to identify the style with which they were designed. Moreover, the inclusion of information from the context of each line shall help improve the predictive capacity of our method.

A first step in the research is the development of the “ground truth” data for our sample, in this case, the OSM lines for the aforementioned parks. We propose a categorisation of parks according to their historical style and then use it to label the lines according to the style with which they were conceived.

(Turner, 2005) proposes 24 historical “styles” that through, a combination of topography, water, paths and vegetation, manage to convey a specific character of a park or designed landscape, covering most design trends applicable to western landscape design. We select from this list those categories applicable to the largest parks in London and add other categories to make them specific to the landscapes of this city:

1- COMMON – HEATH: This “style” is defined by the presence of “desire lines” or natural movement lines in a wild or non constructed environment. These can sometimes be associated with old lines coming from hunting grounds, but most likely these are going to be linked to the movement of people in non-formalised open spaces that are kept in semi-natural conditions. This category is not included in the list by (Turner, 2005) but comprises the largest set of parks in London with most areas of Bushy Park, Richmond and Hampstead falling under this category.

2- PICTURESQUE: These sections of parks correspond to a late romantic approach is characterised by a rejection of axial formality and inclusion of curvy and winding paths avoiding straight visual connections.

3- BAROQUE - STAR: This style is characterised by long straight lines, typically colinear across intersections and forming sharp triangles/polygons or star-shaped patterns, corresponding to what Turner defined as High-Baroque.

5- FORMAL – RENAISSANCE. Smaller-scale rectangular groups and grids, sometimes including rotundas or circular elements (fountain or similar). This includes both landscapes designed in the XVI century as well s Italianate gardens from the XIX century romantic period.

4- PARTERRE: This corresponds to lines around buildings that correspond to circulation, parterres or paved offset from buildings. These lines typically derive strong orthogonality from the architecture as well as a smaller scale and average length. This style was not originally incorporated in (Turner, 2005).

6- CONTEMPORARY FORMAL: These are parks built during the 20th century, mainly focusing on strong or curvy shapes.

For each of the parks in the database, we research information on when it was built or designed and to which style do they belong. This included information about the overall structure as well as smaller internal details or additions. We then carry out labelling of individual lines from all parks, obtaining a database of 10,000 lines which will work as our “ground truth”.

We then define a series of geometrical features used to characterise the lines of the map and which will be used for further correlation with the ground truth data. We call this process Geometric Feature Extraction (GFE). We clean the OS dataset by removing double and triple lines for roads and complex intersections and generate breaking points in 90-degree bends. We then generate the GFE process obtaining 8 measurements for each line as shown in the image below (Figure 1).

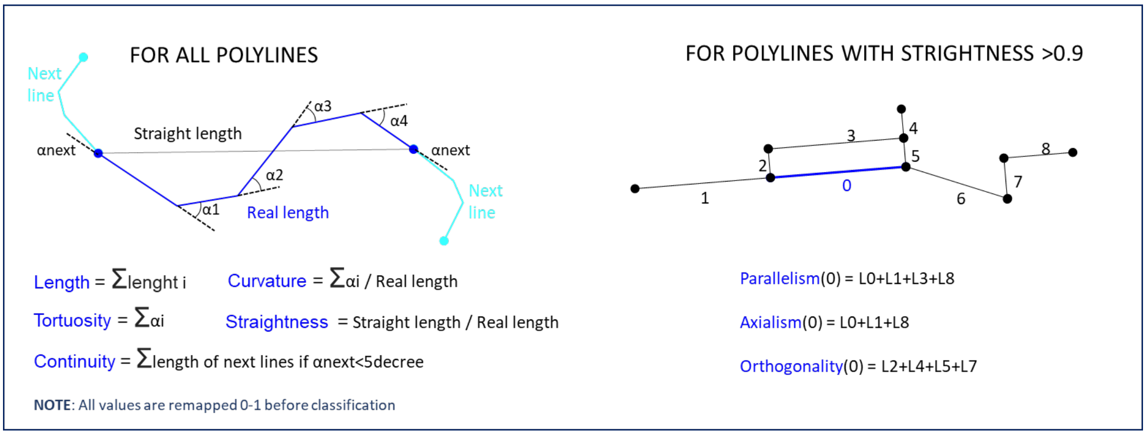

Figure 1 Primary clustering K=5, cutoff depth = 6

Once the initial set of features are calculated for each line, we gather data from the surrounding context by defining a centred sub-graph with a fixed number of topological steps. Data from all these lines in the subgraph is averaged, giving 8 extra features which we call “Average Surrounding” variables. These are remapped 0-1 and concatenated at the end of the previous feature vector for a total of 16 variables for each line of the dataset.

We carry out a brief test to understand the effect of the size of the sub-graph in the overall context definition. The underlying hypothesis is that large numbers of cut-off depth will fail to detect differences in the small part structures since they will fabricate large averages that homogenise data across lines. To do this fine-tuning we perform a K-means clustering on the 16 features of the context-rich extraction to find lines that have some similar grouping or characteristic. We call this step “primary clustering”. We map this clustering onto the site by colour-coding the lines with their cluster number. Some zones of the map show groups of lines with the same colour, indicating that areas with a consistent character. An initial visual assessment seems to indicate that Kmeans of 5 and cut-off distance of 6 topological steps was giving consistent data that could be used in further steps. As it can be seen in Figure 2 larger parks such as Hyde Park are broken down into smaller areas which would have a distinct character such as the Lady Di memorial (green) and st Margaret ward (orange area to the south) without the star-shaped systems above.

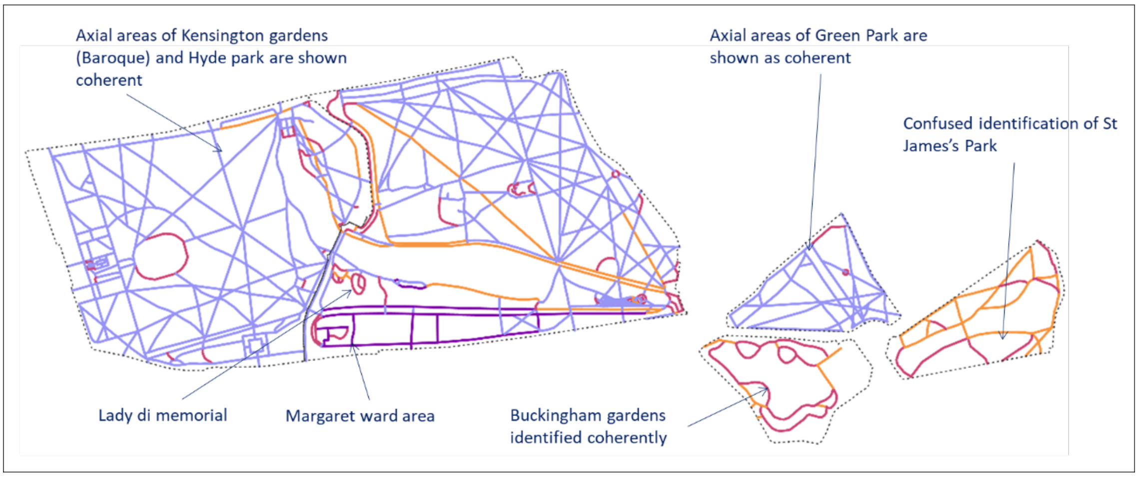

Figure 2 Primary clustering K=5, cutoff depth = 6. Lines coloured according to their "k" cluster

Once we have obtained a rich definition of the context for each line we try to identify groups of lines with a coherent morphological within the parks. The idea is that the identification of the historical character of these structures will be more powerful than that of the single lines since the process of generation of these structures encodes a larger level of context. We call them "parklets" since they may sit within larger areas of a park or be the entire park itself. For each of the primary clusters (k=5), we carry out a point subdivision of the lines in 50m intervals where we run a DBScan to find groups of points with a coherent character that are close to each other. Small spatial clusters smaller are discarded using the *minimum cluster size* variable.

A Convex hull is then developed around each of these spatial clusters which are given the cluster number of the lines used to generate them. This process is shown in Figure 3 where the convex hulls appear as green areas corresponding to the red primary cluster. In principle, since the clusters are defined in space, lines of different primary clusters may be present within that given convex hull. We estimate the proportion of points within the convex hull that coincide with its cluster number and call it *coherence*. The spatial clusters with high coherence (value larger than a given *coherence threshold*) are considered to be valid parklets. For a valid parklet, GFE information is calculated as the average of GFE of all lines inside and the parklet classification historical class with a majority of lines within it.

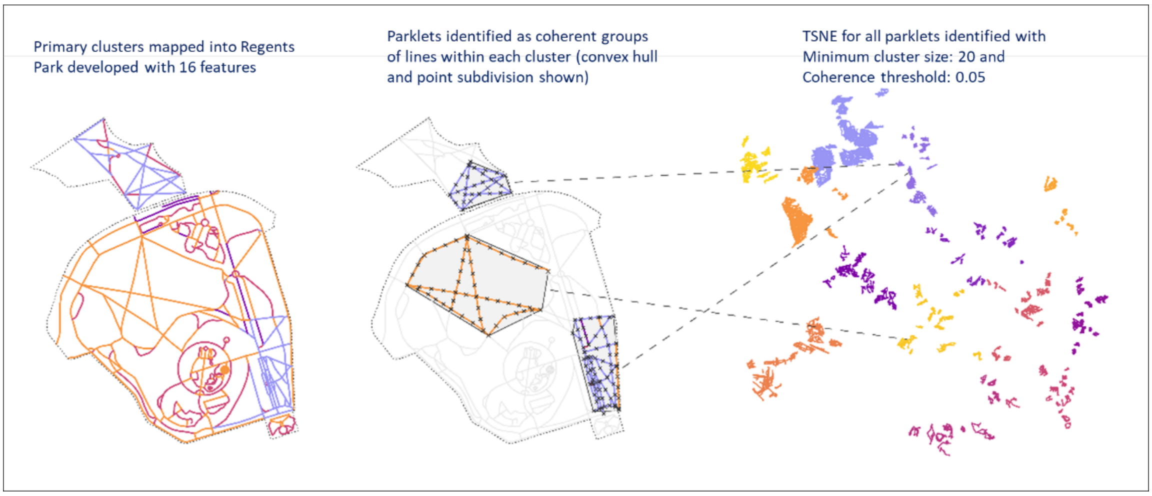

Figure 3 Parklet development and example of TSNE clustering of different parklets

The first two variables that define the nature of the parklets come from the generation of primary clusters. These are the depth of the graph (number of topological steps) and k number of primary clusters. The second variables come from the spatial definition of the parklets and are the minimum cluster size as well as the coherence threshold. We run several tests with these four variables to obtain different groups of parklets using values of depth =[3, 4, 5, 6, 7, 8], k=[4, 5, 6, 7], coherence threshold [5%, 15%, 50%] and minimum cluster size [10, 50, 100] generating a total of 216 sets. Removing duplicate parklets around 8,000 different data points. Figure 3 (right) shows a scatterplot representation of one of these sets where a TSNE is developed on the 16 features, producing a 2D space where the parklets are plotted and colour-coded by proximity. In this test, high values of coherence are likely to reduce the number of clusters since this value will constrain valid item numbers. Low values of minimum size will produce multiple small but coherent pieces of fabric, thus generating a greater number of parklets. A total of 7,500 parklets are produced with the 54 embeddings tests and are used as our last Dataset 3.

After the previous process, we have three datasets that will be used in the following classification step. Dataset 1 (lines with simple GFE) has around 10,000 lines with 8 features and 1 historical category for each line. Dataset 2 (context-rich lines)m has those same lines with 16 features and the same historical category. Dataset 3 (parklets) consists of 8,000 parklets with 16 features each and 1 historical category.

Once we have obtained GFE for the three datasets we use several classifiers to predict the historical categorisation from the features to see the extent to which geometrical information encodes representations that can be traced to its historical character. This process is developed in stages for each of the datasets to understand what are the most representative features, the best classification methods as well as the impact of context in the process. In all classification process, we use a 70%-30% train-test split to develop the classifier and test its performance. We train and test the different classifiers and select the one that reaches the highest accuracy. For the most successful classifier, we apply a variable score assessment and identify the features that carry out most information. We then apply univariate feature selection on the classification process and test the accuracy of the classifier by removing variables progressively starting from the ones with lower score level. In this process, the accuracy should peak when all non-relevant variables have been removed. We identify the peak of the performance as well as an ideal set of variables. We finally develop a confusion matrix for all cases and identify which historical classification styles are better identified in the process.

Experiment result analysis

Classification results for Dataset 1 (GFA for single lines) are relatively poor with a maximum value of 0.558 using Random Forest with all 8 variables ad 0.564 when selecting the 7 most representative variables. The confusion matrix for this latest classifier shows relatively high levels of confusion in most styles, with only "Heath" being properly identified. In this case, parallelism seemed to be the most salient feature to help in the classification (note that this is one of the features that relies slightly on the context).

There is a substantial improvement of accuracy results when we add the first level of context awareness from average GFA values in the sub-graph for each line. Results from Dataset 2 show an optimal result classifier using Nearest Neighbours with a 0.843 accuracy level. Variable scoring indicates a strong increase in the relative value of features including average values for context, with a peak in Average surrounding tortuosity. When applying feature selection for the Nearest Neighbours classifier we find a peak of 0.95 with 8 variables. This indicates that the classifier is discarding almost all non-contextual information as "noise" and only using averages derived from the context as valuable information for the assessment. Using the classification of the parklets gives a 0.93 maximum accuracy for the optimised Nearest Neighbours classifier, suggesting that the selection of context in these items is relatively good.

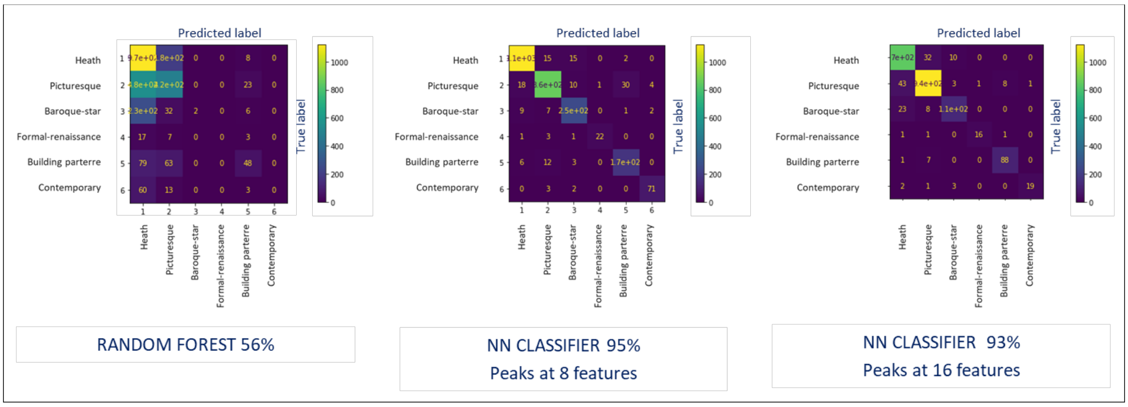

Figure 4 Classification results

Conclusions and further steps

Evidence suggests that analysing single line networks using GFE is an effective mode of encoding geometry, grouping it by a recognisable character and also allowing accurate prediction of landscape character when applied to landscape networks. Increasing the level of context information in the encoding of networks has a positive effect on the prediction of this character. We can also ascertain that these methods also enable the efficient identification of landscape structures that may be hidden or embedded in larger structures, such as the case of our system of definition of parklets.

Regarding line categorization from maps, current ML architectures are being deployed to encode sequential and graph data more efficiently. Further work can be undertaken to encode large sections of the urban fabric and correlate their morphology to various types of data beyond the stylistic. Moreover, methods such as transformers can be deployed to learn from graph datasets and generate new fabric, in this case, incorporating complete structures which incorporate non-homogeneous items in predefined sequences or spatial relations (ie rotunda with axis and adjacent grids etc) which so far are not captured in generative algorithms based on L-systems.

Acknowledgements

This work is part of an EPSRC funded PhD work in collaboration with Space Syntax Limited.

References

1. Boeing, G., 2018. A Multi-Scale Analysis of 27,000 Urban Street Networks: Every US City, Town, Urbanized Area, and Zillow Neighborhood. Environment and Planning B: Urban Analytics and City Science 47, 590–608. https://doi.org/10.1177/2399808318784595
2. Bubalo, M., van Zanten, B.T., Verburg, P.H., 2019. Crowdsourcing geo-information on landscape perceptions and preferences: A review. Landscape and Urban Planning 184, 101–111. https://doi.org/10.1016/j.landurbplan.2019.01.001
3. Dibble, J., Prelorendjos, A., Romice, O., Zanella, M., Strano, E., Pagel, M., Porta, S., 2019. On the origin of spaces: Morphometric foundations of urban form evolution. Environment and Planning B: Urban Analytics and City Science 46, 707–730. https://doi.org/10.1177/2399808317725075
4. Fleischmann, M., 2019. momepy: Urban Morphology Measuring Toolkit. Journal of Open Source Software 4, 1807. https://doi.org/10.21105/joss.01807
5. Golan Levin, 2017. Terrapattern (Overview and Demonstration).
6. Gonzalez, M., Alhasoun, F., 2019. Streetify: Using Street View Imagery And Deep Learning For Urban Streets Development [WWW Document]. GroundAI. URL https://www.groundai.com/project/streetify-using-street-view-imagery-and-deep-learning-for-urban-streets-development/1 (accessed 3.16.21).
7. Hillier, B., Hanson, J., 1984. The Social Logic of Space. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511597237
8. Law, S., Berghauser Pont, M., Shen, Y., Penn, A., 2019a. Identifying Street-Character-Weighted Local Area using locally weighted community detection methods. The case study of London and Amsterdam.
9. Law, S., Paige, B., Russell, C., 2019b. Take a Look Around: Using Street View and Satellite Images to Estimate House Prices. ACM Trans. Intell. Syst. Technol. 10, 1–19. https://doi.org/10.1145/3342240
10. Law, S., Paige, B., Russell, C., 2018. Research on Visual Desirability [WWW Document]. URL https://visual-desirability-v3.firebaseapp.com/#/street-view (accessed 3.16.21).
11. Moosavi, V., 2017. Urban morphology meets deep learning: Exploring urban forms in one million cities, town and villages across the planet. arXiv:1709.02939 [cs].
12. Netto, V.M., Brigatti, E., Cacholas, C., 2020. From form to information: Analysing built environments in different spatial cultures. arXiv:2006.13897 [physics].
13. Nice, K.A., Thompson, J., Wijnands, J.S., Aschwanden, G.D.P.A., Stevenson, M., 2020. The "Paris-end" of town? Urban typology through machine learning. Urban Science 4, 27. https://doi.org/10.3390/urbansci4020027
14. Oliveira, V., 2019. An historico-geographical theory of urban form. Journal of Urbanism: International Research on Placemaking and Urban Sustainability 12, 412–432. https://doi.org/10.1080/17549175.2019.1626266
15. Redi, M., Aiello, L.M., Schifanella, R., Quercia, D., 2018. The Spirit of the City: Using Social Media to Capture Neighborhood Ambiance. Proc. ACM Hum.-Comput. Interact. 2, 144:1-144:18. https://doi.org/10.1145/3274413
16. Scepanovic, S., Joglekar, S., Law, S., Quercia, D., 2021. Jane Jacobs in the Sky: Predicting Urban Vitality with Open Satellite Data. ArXiv.
17. Schirmer, P., Axhausen, K., 2015. A multiscale classification of urban morphology. The journal of transport and land use 9, 101–130.
18. Streetscore, 2018. Streetscore [WWW Document].
19. Turner, T., 2005. Garden History Philosophy and Design 2000 BC to 2000 AD. Spon Press, London.