"Learning from Deep Learning" Lessons from using computer vision to identify (urban) form and function in open data satellite imagery

Martin Fleischmann[*][†][§]
Daniel Arribas-Bel[*][†][‡][¶]
Geographic Data Science Lab, University of Liverpool
The Alan Turing Institute

Abstract: The building blocks that make up cities -the activities and agents conceptualised as urban function, and the structure that supports them conceptualised as urban form- can be spatially arranged in many ways. This paper relies on the concept of "spatial signatures", a characterisation of space designed to understand urban environments, that exhaustively divide geographical space into distinct classes based on form and function. Due to the dependency on data sources that are being updated at a variable rate, signatures cannot be easily updated with frequency. One possible solution comes from remote sensing and satellite imagery. While staying in the realm of open data, we explore this pathway using the Sentinel-2 imagery within a deep convolutional neural network (CNN) trained to predict spatial signature type across Great Britain. Our focus is not only to develop a performant CNN but also to learn about the nature of the classes we try to predict (appropriate scale, inter-class relationships) through the lens provided by the neural network. With Sentinel-2 being relatively coarse in the resolution, there are not only technical questions of the CNN architecture, but also geographical ones related to the Modifiable Areal Unit Problem and the ability of samples to capture the nature of each signature type. Furthermore, there is the question of to which degree signatures can be seen from space. We present exploratory work and empirical experiments, and discuss the opportunities and challenges in using remote sensing to reliably detect concepts like spatial signatures using openly available satellite imagery.

Key words: spatial signatures, classification, remote sensing, artificial intelligence, open data

[*] THANKS.

[†] Geographic Data Science Lab, Department of Geography and Planning, University of Liverpool, Roxby Building , 74 Bedford St S , Liverpool , L69 7ZT, United Kingdom

[§] E-mail: M.Fleischmann@liverpool.ac.uk; website: https://martinfleischmann.net/.

[‡] E-mail: D.Arribas-Bel@liverpool.ac.uk; phone: +44 (0)151 795 9727; website: http://darribas.org.

[¶] The Alan Turing Institute, British Library, 96 Euston Road, London, England, NW1 2DB, United Kingdom

1. Introduction

2. Materials and Methods

In this section we present the materials used in the research - the British spatial signatures used as to generate labels for individual chips and Sentinel 2 satellite imagery - and methods designed to unpack the role of geography in image-based deep learning.

2.1 Materials

The whole research uses only two data inputs, one representing the "ground truth" we are aiming to predict using neural networks and the other representing the satellite imagery. While the latter does not need a lot of introduction, the British spatial signatures used as the labels need to be explained further.

2.1.1 British Spatial Signatures

Spatial signatures are a way of classification of space covering entirety of a case study area. They are defined as *"a characterisation of space based on form and function designed to understand urban environments"* (Arribas-Bel and Fleischmann, 2022) and the definition already points at the clear distinction between signatures and traditional Land Use / Land Cover (LULC) classifications. Taking the example of CORINE (REF) as a representative of LULC, it has 44 distinct classes, out of which 2 cover urban form and other 6 can be loosely related to urban areas[1]. A similar situation is with recently released global LULC datasets. European Space Agency's WorldCover project distinguishes 11 classes of which one is urban (Built-up) REF. Esri's Land cover has 9 classes where one is *Built Area* and the rest covers unbuilt areas REF. This ratio of built vs unbuilt classes is typical but not very suited for research applications focusing on urban environments. Spatial signatures tend to flip this ratio as they are primarily classifying urban space.

There are two key main concepts embedded in spatial signatures delivering urban-focused classification. The first one is the spatial unit called the enclosed tessellation cell (ETC). To derive ETCs, we first generate *enclosures*, a space fully enclosed by a set of barriers (roads, railways, rivers, coastline). ETCs are then a result of Voronoi tessellation based on building footprint polygons. The resulting spatial unit has adaptive granularity reflecting the scale of each individual urban pattern. The second is the selection of characters describing each ETC. We measure form and function, both primarily urban phenomena and mostly omit environmental aspects focusing on land cover patterns.

British spatial signatures as presented in Fleischmann and Arribas-Bel (2022) are one application of the concept of spatial signatures in the context of Great Britain. It divides the space into 16 data-driven classes (Figure 1) listed in Table 1. Out of these 16 classes, nine are entirely urban, four are peripheral and only three are classifying natural spaces, inverting the ratio of built vs unbuilt classes known from LULC. However, out of these 16 classes, some are very rare and it

[1] Continuous urban fabric, Discontinuous urban fabric; Construction sites, Green urban areas, Sport and leisure facilities, Industrial or commercial units, Road and rail networks and associated land, Port areas

	total area (sq.km)	total ETC count	percentage of area	percentag
signature_type				
Countryside agriculture	93,856.1	3,022,385	41	
Accessible suburbia	2,244.5	1,962,830	1	
Dense residential neighbourhoods	957.2	502,835	0	
Connected residential neighbourhoods	565.4	374,090	0	
Dense urban neighbourhoods	570.6	238,639	0	
Open sprawl	5,081.5	2,561,211	2	
Wild countryside	91,306.3	595,902	40	
Warehouse/Park land	2,462.4	707,211	1	
Gridded residential quarters	261.2	209,959	0	
Urban buffer	31,588.8	3,686,554	14	
Disconnected suburbia	708.9	564,318	0	
Local urbanity	231.1	86,380	0	
Concentrated urbanity	7.8	1,390	0	
Regional urbanity	76.4	21,760	0	
Metropolitan urbanity	16.5	3,739	0	
Hyper concentrated urbanity	2.2	264	0	

Table 1: Classes of British spatial signatures and their coverage in terms of area and a number of ETCs.

would not be feasible to attempt to predict them. Therefore, we merge five classes falling under the "urbanity" group into a single one and use resulting 12 classes throughout this paper.

2.1.2 Sentinel 2 imagery

The second data input used in this research is satellite imagery provided by the Sentinel 2 mission. Specifically, we use the pre-processed cloud-free mosaic of Sentinel 2 released by Corbane et al. (2020). The mosaic provides pixel-level composite based on imagery for the period January 2017-December 2018 at an original resolution of 10 meters per pixel. While Sentinel 2 captures many spectral bands beyond traditional visible red, green and blue (RGB), this research uses only RGB bands due to its employment of pre-trained neural networks stemming from non-satellite imagery that is composed only of RGB. The exclusion of other bands may be seen as a limiting factor of the work, but we believe that, as with other aspects that will be discussed later, it efficiently illustrates the *lower bound* of the performance of presented method and can be only improved with addition of further spectral bands or other data (e.g. synthetic-aperture radar imagery).

Another notable aspect of the Sentinel 2 imagery is the resolution. Ten meters per pixel may be enough to distinguish LULC classes as shown by the research project discussed above. However, there is the question of whether it is enough to segment urban environments. Individual buildings often do not stretch beyond the spatial extent of two pixels, which is severely limiting what we can *see* on the image, as illustrated in Figure 1. While other data sources may provide better resolution (REFS), potentially improving model performance, this research is bound within the limits of *open data*, where Sentinel 2 is the best offering to date.

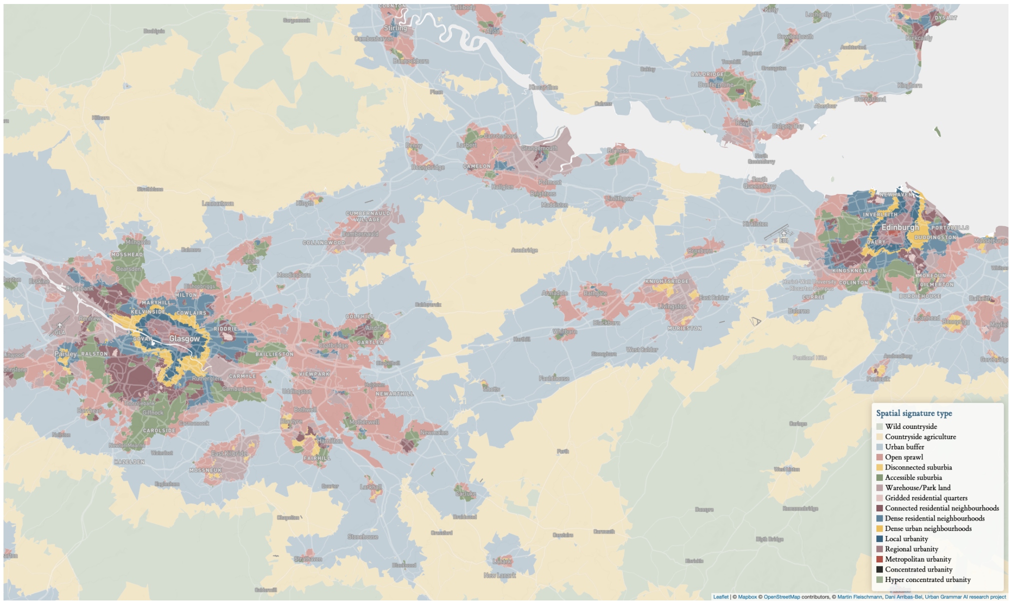
Glasgow
Edinburgh
Spatial signature type
Wild countryside
Countryside agriculture
Urban buffer
Open sprawl
Disconnected suburbia
Accessible suburbia
Warehouse/Park land
Gridded residential quarters
Connected residential neighbourhoods
Dense residential neighbourhoods
Dense urban neighbourhoods
Local urbanity
Regional urbanity
Metropolitan urbanity
Concentrated urbanity
Hyper concentrated urbanity
Leaflet | © Mapbox © OpenStreetMap contributors, © Martin Fleischmann, Dani Arribas-Bel, Urban Grammar AI research project

Figure 1: Spatial signatures in the area of the Scottish Central Belt stretching from Glasgow to Edinburgh.

2.2 *Methods*

We define our challenge as an image classification task and use competing alternatives to explore which one performs best. Each of them imply geographically relevant trade-offs. First, we build and train a model composed of a convolutional neural network and probability modelling able to predict the 12 classes derived from the spatial signatures. Second, we use methods designed to unveil which of the inherently geographical decisions that are being tested has significant effect on the resulting performance and should therefore be taken into account when applying CNN on spatial problems.

Overall, our exercise is structured as a comparison of models that attempt to predict the 12 spatial signatures entirely from Sentinel 2 imagery. Each model takes a set of chips as input, runs the class prediction using the convolutional neural network (CNN) and builds a (spatial) model on top of the resulting probabilities. The differences between the models are capturing the geographical options being tested - an extent of the area sampled from the satellite imagery into a single *chip*, presence of spatial augmentation, class exclusivity within each chip, and an architecture of probability modelling on top a prediction coming from the CNN.

Finally, performance of each model is assessed using both traditional non-spatial techniques used in deep learning and bespoke spatial metrics. Given the large number of resulting values, a regression approach is used to determine the effect of the tested options.

Each of the steps is further discussed in detail in the subsequent sections.

2.2.1 *Chip size*

The first question that needs to be answered when trying to apply a classification algorithm on a contiguous satellite imagery is how to sample such data into individual chips that can be assigned

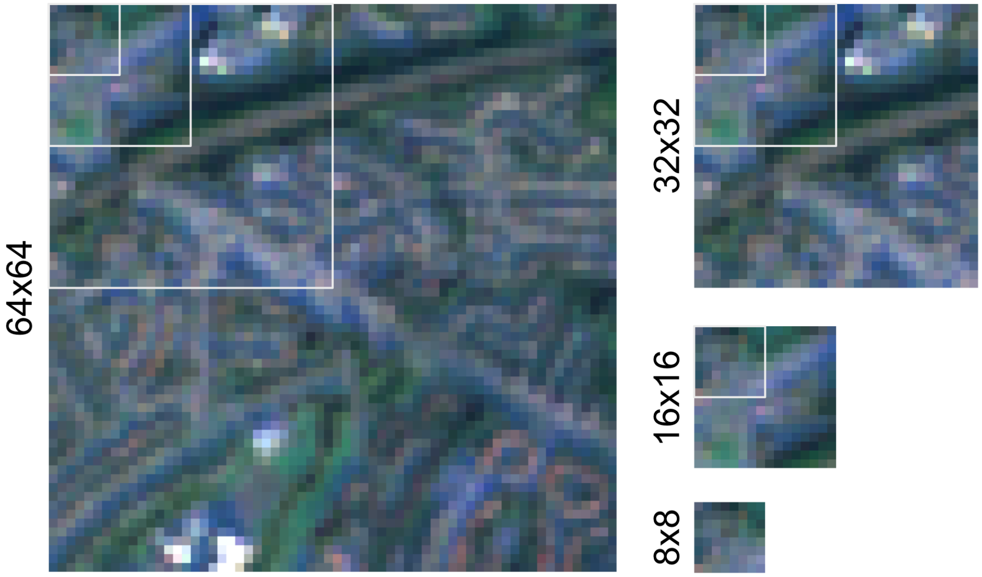

Figure 2: Illustration of the selected chip sizes using the Sentinel 2 cloud-free mosaic. Each of the chips also shows the sizes of the smaller options as a white outline.

to classes. Pre-trained CNNs usually expect a square image of a certain size but that does not mean that the same size (in terms of pixels) needs to be directly sampled from the image thanks to possible resampling. What should be retained though, is the ratio. Therefore, we need to sample square chips of of a relatively custom size. Within image classification problems, we should also assume that each chip contains data of a single class only, therefore, such a chip should be fully within a boundary of a single signature type. That poses some restrictions as spatial signatures, especially in urban context, tend to be relatively granular and large chips would simply not fit inside the boundaries. The goal is therefore to find a balance between the number of chips that can be sampled from the data and an amount of information each chip can hold. Given the relatively coarse resolution of Sentinel 2, a chip of 100x100 meters consists of only 10x10 pixels, which may not be enough to capture the nature of a signature type and distinguish it from other types. On the other hand, a chip of 1000x1000 meters, that is likely large enough to capture the difference, will not fit in most of the signature boundaries and we would end with only a few chips per urban class.

Literature very rarely discusses the decision-making process when defining the chip size. In some cases, the size is predetermined due to the requirement of either a pre-trained model REF or existing set of labelled data REF. In other, the size that has been used in other studies is simply applied again without discussing the implications of such a decision REF. That is surprising as the chip size is a prime example of the Openshaw effect (also known as MAUP, REF), especially its scale part, which states that a change of the scale may affect the outcome of an experiment. Hence such an effect should be at least considered in an interpretation if not intentionally minimized.

In this work we try to understand the effect of a chip size by testing all the models based on four different chip sizes - 80, 160, 320 and 640 meters representing chips of 8x8, 16x16, 32x32 and 64x64 pixels, illustrated on a Figure 1.

2.2.2 *Data (spatial) augmentation*

As mentioned above, certain chip sizes, in combination with the signature geometry and a requirement to keep them exclusively within a single class, may result in an insufficient amount of training data for some signature types. Under-sampling like this one can be a serious problem, that is not unique to spatial modelling. However, the traditional augmentation methods are not entirely applicable here. For example, in an image classification problem trying to determine if there is a cat or a dog on an image, we can flip the picture along y axis, add some rotation or zoom to get more versions of the same image and expand the set of training data. Neither of these methods is applicable to the spatial problems. Flipping or rotating the image would break the natural light conditions, while zooming in would change the scale of urban environment we are attempting to capture. On the other hand, the geographical nature of the problem allows us to use spatial augmentation technique we call *sliding*.

Sliding can be seen as overlapping sampling. Instead of overlaying a grid of chips over target geometry and using each pixel only once, we take the initial grid and slide it a few pixels horizontally and vertically as illustrated on a Figure 3. If the boundary of a slid chip is fully within a signature geometry, it is added to the pool of chips to be used. This process is done repeatedly to ensure that each class has a reasonable amount of chips to work with.

It is to be noted, that sliding can cause a data leakage (sequences of pixel being present in both training and validation sets) if done before splitting the data into training and validation subsets. Therefore, we first create the initial grid, subdivide it spatially into four parts (40% for CNN training, 10% for CNN validation, 40% for probability modelling training, 10% for probability modelling validation) and apply sliding within each part to avoid any pixels being shared among chips from different sets.

2.2.3 *Model architecture*

Architecture of each model is composed of two parts. First is the CNN predicting a probability that a chip belongs to a class (or a proportion of a chip covered by a class). Second is the machine learning modelling, taking the output of the CNN and predicting the dominant class. This section describes each of these in detail. It is to be noted, that contrary to the majority of deep learning-focused research articles, the actual architecture of CNN is not of a particular interest in this paper. It is assumed that the effect of geographic decisions will show similar behavior no matter the architecture (within some limits). This work therefore uses EfficientNetB4 pre-trained on ImageNet dataset. An appendix A. shows a brief comparison of several standard neural network architectures and their performance on a subset of data to illustrate the point. The top layer of the pre-trained model is then replaced by a custom sequence of dense layers described below.

The default approach is a standard image classification problem, using the sets of chips that are fully within a single signature types, both with and without augmentation using sliding. The custom top layer of the pre-trained CNN then contains a Global Average Pooling (2D) layer, a dense layer with ReLu activation and 256 neurons, and a dense layer with the softmax activation

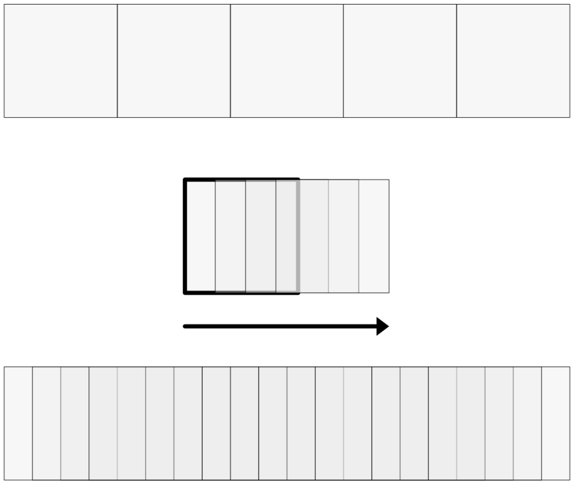

Figure 3: Diagram illustrating the sliding mechanism in one direction. The first row shows the initial non-overlapping grid, the last one final overlapping set of chips. The same approach is then also applied vertically.

and number of neurons equal to a number of classes (12). A result for a single chip is a an array of a length 12 containing probabilities that a chip belongs to each class. The sum of all probabilities is 1.

However, there is one more option how to tackle the problem, apart from the image classification. When we relax the requirement that every chip must be fully within boundaries of a single signature type, we end up with chips encoding proportions of area belonging to each class. Instead of a single label per chip, we now deal with an array. This can be beneficial from the geographical perspective as such chips now inherently encode co-location of individual signature types and a model can use this information during the prediction. As signature types usually tend to neighbor only a subset of other classes (e.g., Urbanity never neighbors Wild Countryside), we can assume that an information on co-location can positively impact the resulting performance. For these reasons, we include a set of chips sampled from a grid crossing the boundaries of signature types (using the same chip sizes as before) and adapt the CNN for the multi-output regression problem instead of image classification one. That means that the top layer is composed of a Global Average Pooling (2D) layer a dense layer with ReLu activation and 256 neurons and a dense layer with the sigmoid activation and a number of neurons equal to a number of classes (12). The result for a single chip is a similar array but containing predicted proportions. The sum of all proportions can be anything within 0 and 12x1.

[Spatial modelling of probs. DAB to do]

2.2.4 *Performance metrics*

[DAB to do]

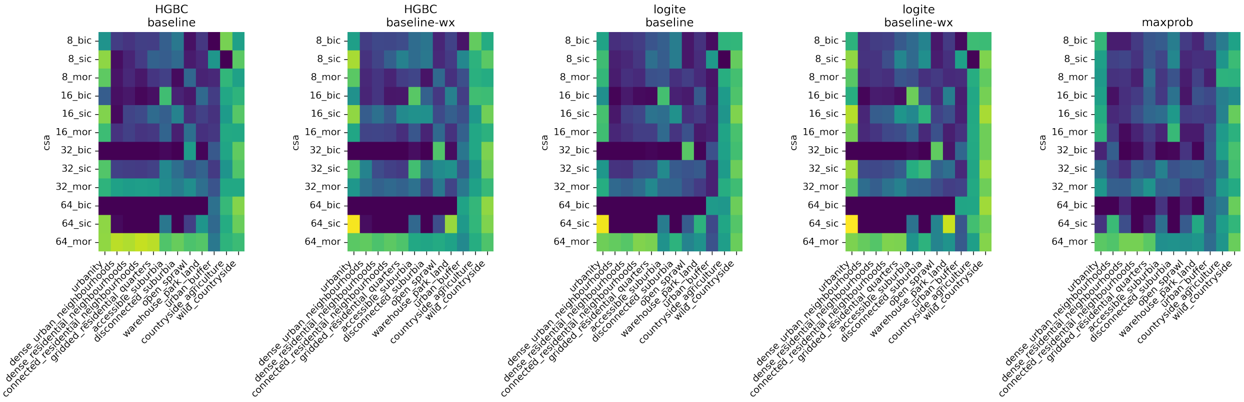

Figure 4: Within-class accuracy scores grouped by model. Each panel represents results from one of the five models compared, namely: histogram-based boosted classifier (`HGBC`) with features pertaining only to a given chip (`baseline`) or including also features from neighbouring ones (`baseline-wx`); Logit ensemble (`logite`) with the same two variations; and a simpler maximum probability approach (`maxprob`). Each row in the heatmap corresponds to a pair of chipsize (8, 16, 32, and 64 pixels) and architechture (baseline image classification, or `bic`; sliding image classification, or `sic`; and multi-output regression, or `mor`) used in the neural network stage of the pipeline. Colouring is standardised across panels and values range from 0 (dark purple) to 1 (bright yellow).

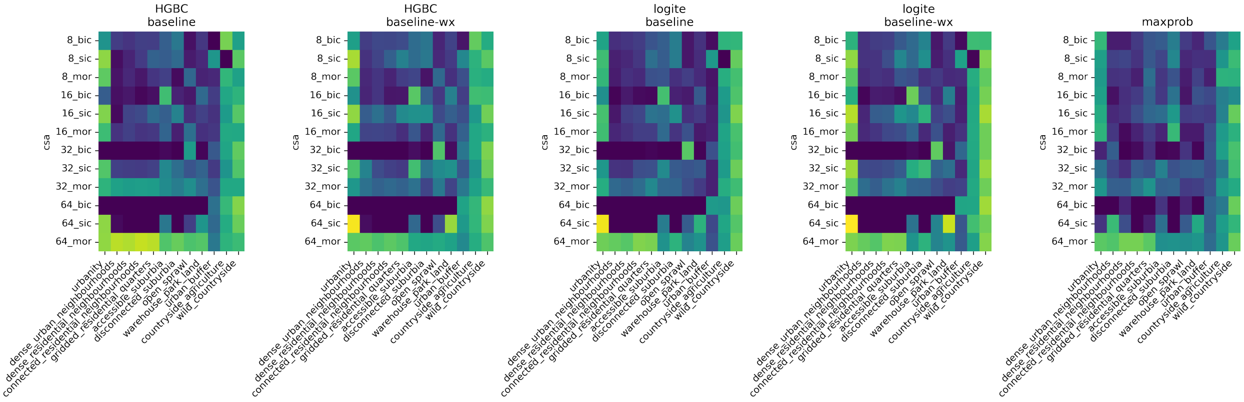

Figure 5: TBC

2.2.5 *Summarizing experiments*

[DAB to do]

3. Results

Within-class accuracy by model can be seen in Figure 4 (a sister figure where scores are grouped by signature rather than by model can be found in Appendix B.).

Appendix C. shows confusion matrices.

4. Discussion

	κ	Global Accuracy	Macro F1 w.	Macro F1 avg.
Intercept	0.2185***	0.3236***	0.2790***	0.1798***
	(0.0209)	(0.0175)	(0.0174)	(0.0375)
(M) Logit E.	-0.0245	-0.0256*	-0.0324**	-0.0325
	(0.0168)	(0.0141)	(0.0141)	(0.0302)
(M) Max. Prob.	-0.0559**	-0.0606***	-0.0421**	-0.0296
	(0.0222)	(0.0187)	(0.0186)	(0.0399)
(A) M.O.R.	0.0227	-0.0357**	-0.0278*	0.1787***
	(0.0184)	(0.0155)	(0.0154)	(0.0331)
(A) S.I.C.	0.0232	-0.0247	-0.0171	0.1101***
	(0.0184)	(0.0155)	(0.0154)	(0.0331)
Chip Size	0.0036***	0.0043***	0.0048***	0.0014**
	(0.0004)	(0.0003)	(0.0003)	(0.0006)
W	0.0572***	0.0468***	0.0531***	0.0392
	(0.0168)	(0.0141)	(0.0141)	(0.0302)
R^2	0.7214	0.8281	0.8514	0.4191
R^2 Adj.	0.6899	0.8086	0.8346	0.3533
N.	60	60	60	60

Table 2: Regression outputs explaining global non-spatial performance scores. Explanatory variables with a preceding (M) and (A) correspond to binary variables for the type of model (with histogram-based boosted classifier, or HGBC, as the baseline) and architecture (with baseline image classification, or BIC, as the baseline), respectively. Standard errors in parenthesis. Coefficients significant at the 1%, 5%, 10% level are noted with ***, **, and *, respectively.

	Within-Class Accuracy		
Intercept	0.1866***	-0.0237	0.0595**
	(0.0308)	(0.0311)	(0.0303)
(M) Logit E.	-0.0125	-0.0125	-0.0125
	(0.0159)	(0.0141)	(0.0146)
(M) Max. Prob.	-0.0188	-0.0188	-0.0188
	(0.0211)	(0.0186)	(0.0193)
(A) M.O.R.	0.1753***	0.2512***	0.1753***
	(0.0175)	(0.0163)	(0.0160)
(A) S.I.C.	0.1202***	-0.0783***	0.1202***
	(0.0175)	(0.0209)	(0.0160)
Chip Size	0.0014***	0.0041***	0.0014***
	(0.0003)	(0.0003)	(0.0003)
1k Obs.		0.0514***	
		(0.0036)	
% Obs.			0.0156***
			(0.0013)
W	0.0365**	0.0365***	0.0365**
	(0.0159)	(0.0141)	(0.0146)
(S)Urbanity	0.2358***	0.2022***	0.2574***
	(0.0349)	(0.0309)	(0.0320)
(S)Dense urban neighbourhoods	-0.1420***	-0.1075***	-0.0998***
	(0.0349)	(0.0309)	(0.0322)
(S)Dense residential neighbourhoods	-0.1414***	-0.0836***	-0.0983***
	(0.0349)	(0.0311)	(0.0322)
(S)Connected residential neighbourhoods	-0.1306***	-0.0726**	-0.0754**
	(0.0349)	(0.0311)	(0.0323)
(S)Gridded residential quarters	-0.0785**	-0.0127	-0.0049
	(0.0349)	(0.0312)	(0.0326)
(S)Disconnected suburbia	-0.0601*	-0.0103	-0.0019
	(0.0349)	(0.0311)	(0.0324)
(S)Open sprawl	-0.0845**	-0.0995***	-0.1143***
	(0.0349)	(0.0309)	(0.0321)
(S)Warehouse park land	-0.0857**	-0.0788**	-0.0817**
	(0.0349)	(0.0309)	(0.0320)
(S)Urban buffer	-0.0828**	-0.1382***	-0.1753***
	(0.0349)	(0.0311)	(0.0330)
(S)Countryside agriculture	0.2236***	0.1593***	0.1118***
	(0.0349)	(0.0312)	(0.0334)
(S)Wild countryside	0.3876***	0.3283***	0.2925***
	(0.0349)	(0.0311)	(0.0330)
R^2	0.4979	0.6087	0.5794
R^2 Adj.	0.4857	0.5987	0.5686
N.	720	720	720

Table 3: Regression outputs explaining within-class accuracy. Explanatory variables with a preceding (M), (A) and (S) correspond to binary variables for the type of model (with histogram-based boosted classifier, or HGBC, as the baseline), architecture (with baseline image classification, or BIC, as the baseline) and spatial signature (with Accessible suburbia as the baseline), respectively. Standard errors in parenthesis. Coefficients significant at the 1%, 5%, 10% level are noted with ***, **, and *, respectively.

	JC	$\log(JC)$	JC	$\log(JC)$
	W_thr	*W_thr*	*W_union*	*W_union*
Intercept	4.3454***	1.4617***	4.7103***	1.6311***
	(0.9507)	(0.1344)	(0.5763)	(0.1080)
(M) Logit E.	-0.1406	-0.0431	0.1851	0.0481
	(0.4951)	(0.0700)	(0.2995)	(0.0561)
(M) Max. Prob.	0.1128	-0.1223	0.2819	0.0223
	(0.6442)	(0.0911)	(0.3887)	(0.0728)
(A) M.O.R.	-3.1630***	-0.5744***	-2.7875***	-0.4647***
	(0.5494)	(0.0777)	(0.3301)	(0.0619)
(A) S.I.C.	0.0119	-0.2390***	-0.6666**	-0.0481
	(0.5532)	(0.0782)	(0.3329)	(0.0624)
Chip Size	0.0297***	-0.0005	-0.0061	-0.0080***
	(0.0108)	(0.0015)	(0.0065)	(0.0012)
W	-0.9325*	-0.1376**	-0.9556***	-0.1785***
	(0.4945)	(0.0699)	(0.2991)	(0.0560)
(S)Urbanity	4.6650***	0.6574***	0.1156	-0.1258
	(1.0696)	(0.1512)	(0.6460)	(0.1211)
(S)Dense urban neighbourhoods	1.7796*	0.5094***	0.7480	0.1609
	(1.0695)	(0.1512)	(0.6487)	(0.1216)
(S)Dense residential neighbourhoods	-0.8545	0.0672	-0.4636	-0.0920
	(1.0958)	(0.1550)	(0.6647)	(0.1246)
(S)Connected residential neighbourhoods	-0.3656	0.1543	-0.4388	-0.1447
	(1.1018)	(0.1558)	(0.6647)	(0.1246)
(S)Gridded residential quarters	-0.2000	0.1009	-0.6203	-0.2111*
	(1.0744)	(0.1519)	(0.6517)	(0.1221)
(S)Disconnected suburbia	-0.9752	-0.1719	-1.0303	-0.3358***
	(1.1213)	(0.1586)	(0.6684)	(0.1252)
(S)Open sprawl	1.8342*	0.1734	2.1575***	0.3576***
	(1.0604)	(0.1499)	(0.6432)	(0.1205)
(S)Warehouse park land	0.5496	0.2123	1.2245*	0.3054**
	(1.0694)	(0.1512)	(0.6487)	(0.1216)
(S)Urban buffer	-0.0558	-0.0931	2.7027***	0.5164***
	(1.0521)	(0.1488)	(0.6382)	(0.1196)
(S)Countryside agriculture	-1.3759	-0.2511*	0.6623	0.0670
	(1.0521)	(0.1488)	(0.6382)	(0.1196)
(S)Wild countryside	-2.0183*	-0.5065***	-0.5918	-0.1635
	(1.0521)	(0.1488)	(0.6382)	(0.1196)
R^2	0.1589	0.1954	0.2118	0.2660
R^2 Adj.	0.1368	0.1743	0.1913	0.2468
N.	665	665	670	670

Table 4: Regression outputs explaining (log of) differences in the spatial pattern between observed and predicted values, as measured by the Join Counts statistic. The Join Counts for each signature were computed using two types of spatial weights: one based on a distance threshold of 1Km (*W_thr*), and another one built as a the union of nearest neighbor and queen contiguity matrices (*W_union*). Explanatory variables with a preceding (M), (A) and (S) correspond to binary variables for the type of model (with histogram-based boosted classifier, or HGBC, as the baseline), architecture (with baseline image classification, or BIC, as the baseline) and spatial signature (with Accessible suburbia as the baseline), respectively. Standard errors in parenthesis. Coefficients significant at the 1%, 5%, 10% level are noted with ***, **, and *, respectively.

References

Arribas-Bel, D. and Fleischmann, M. (2022). Spatial Signatures - Understanding (urban) spaces through form and function. *Habitat International*, 0(0):0.

Corbane, C., Politis, P., Kempeneers, P., Simonetti, D., Soille, P., Burger, A., Pesaresi, M., Sabo, F., Syrris, V., and Kemper, T. (2020). A global cloud free pixel- based image composite from sentinel-2 data. *Data in Brief*, 31:105737.

Fleischmann, M. and Arribas-Bel, D. (2022). Geographical characterisation of british urban form and function using the spatial signatures framework. *Scientific Data*, 9(1):1–15.

Appendix A. Technical appendix

A. Comparison of neural network architechture

B. *Within-class performance by spatial signature*

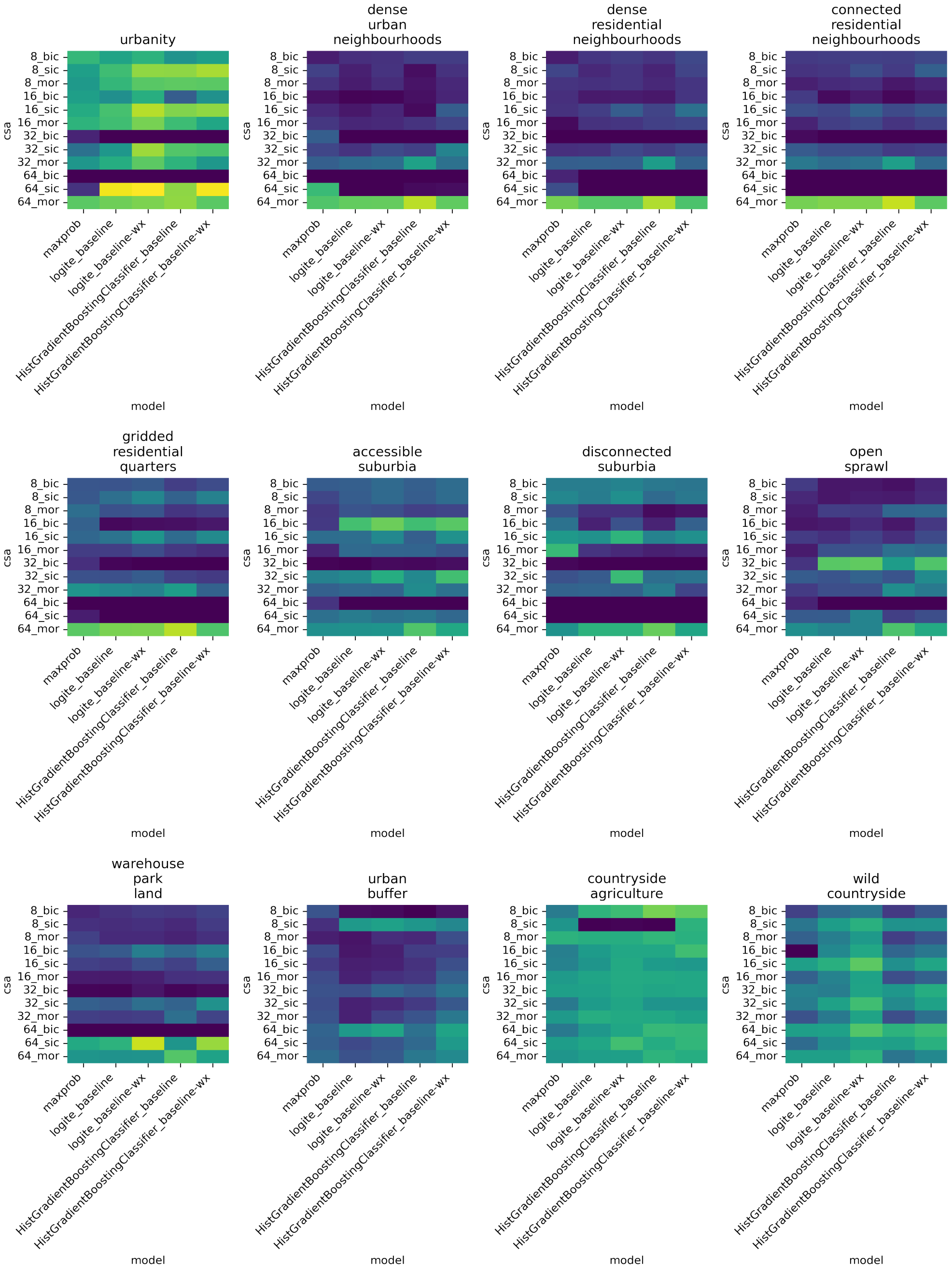

Figure 6: Within-class accuracy scores grouped by signature. Each panel represents results from one of the 12 signatures predicted. Each column in the heatmap corresponds to one of the five models compared, namely: histogram-based boosted classifier (HGBC) with features pertaining only to a given chip (`baseline`) or including also features from neighbouring ones (`baseline-wx`); Logit ensemble (`logite`) with the same two variations; and a simpler maximum probability approach (`maxprob`). Each row corresponds to a pair of chipsize (8, 16, 32, and 64 pixels) and architechture (baseline image classification, or `bic`; sliding image classification, or `sic`; and multi-output regression, or `mor`) used in the neural network stage of the pipeline.

C. Confusion matrices