

A comprehensive vision-based framework for understanding nation-wide urban socio-demographics

Kun Zhou^{a,b}, Yang Chen^{b,c}, Pengyuan Liu^{d,e}, Yujun Hou^{b,f}, Filip Biljecki^{b,g,*}

^a*Department of Geography, National University of Singapore, Singapore*

^b*Department of Architecture, National University of Singapore, Singapore*

^c*School of Geography, Nanjing Normal University, China*

^d*Division of Urban Studies and Social Policy, School of Social & Political Sciences, University of Glasgow, United Kingdom*

^e*Urban Big Data Centre, School of Social and Political Sciences, University of Glasgow, United Kingdom*

^f*Department of Geography and Planning, University of Liverpool, Roxby Building, Liverpool, L69 7ZT, United Kingdom*

^g*Department of Real Estate, National University of Singapore, Singapore*

Abstract

Understanding geodemographic distribution across neighbourhoods is essential for identifying socio-spatial differentiation in cities. Traditional geodemographic classification studies rely primarily on various census and socioeconomic data, offering limited insight into the street-level environments through which such differences are experienced and reproduced. In particular, the visual and perceptual dimensions of neighbourhoods remain largely absent from existing classification frameworks. This study advances geodemographic classification research by introducing street view imagery (SVI) as a novel source for capturing latent visual features linked to socio-spatial structure. We propose an automated machine learning framework that leverages 556 image-derived variables to predict socio-demographic groups. The framework is applied in ten UK cities, where a national model and local instances are developed to assess the extent to which visual-demographic relationships are place-dependent or generalizable. Results show that SVI has potential for capturing geodemographic patterns and enabling national-scale generalization. The local model achieves an accuracy ranging from 0.49 to 0.70, while the national model brings an accuracy of 0.62. Finally, cross-city transfer analysis reveals both shared and heterogeneous visual–demographic relationships across urban contexts. This study contributes a new street-level observational layer that complements traditional data sources and extends the methodological scope of geodemographics, while informing improvements to public spaces and supports policy objectives aimed at fostering inclusive neighbourhood environments.

Keywords: Geodemographic classification; Cross-city analysis; Urban visual analysis; Population mapping

1. Introduction

Social organization of urban space exhibits intricate spatial patterns, specifically in the distribution of population groups (Hu et al., 2024). Understanding these patterns enhances comprehension of the spatial structure of urban society and reveals residential differentiation, social segregation, and inequality across neighbourhoods and cities (Wu et al., 2014; Bruch, 2014). This knowledge also provides essential information for public service allocation, community renewal, and equity assessment in urban planning (Song et al., 2024; Du et al., 2023; Cheng et al., 2022b). Given the significance of these patterns and their variability across urban contexts, more comprehensive and generalizable approaches are needed to characterize the spatial distribution of population groups.

Existing studies have primarily approached the distribution of population groups through geodemographic classification systems, which segment small-area geographies into clusters based on demographic, socioeconomic and other characteristics (Lin, 2024; Singleton and Spielman, 2026). These systems are widely used to reveal neighbourhood differentiation and the socio-spatial organization of urban populations (Lin, 2024). For example, Wyszomierski et al. (2024) utilizes census data to divide geographic areas into groups with similar characteristics, enabling the construction of urban profiles such as student-oriented or retired professionals. With the increasing availability of diverse data sources, existing studies have extended geodemographic classification beyond census data by incorporating other data sources, such as consumer data (Singleton et al., 2020) and financial data (Ye and Singleton, 2026). While these data capture behavioural and socioeconomic correlates of neighbourhoods, they provide limited information on the visible street-level built environments associated with different neighbourhoods. Since living environments form the context in which residents interact and experience daily life, they may offer a complementary perspective for geodemographic classification. This is also consistent with recent calls for richer multi-source data to enhance the learning of geodemographic characteristics (Singleton and Spielman, 2026), as well as ongoing efforts by UK statistical authorities to broaden the use of alternative data sources and provide more frequent and timely population statistics (UK Statistics Authority, 2025; Office for National Statistics, 2026). However, it is not yet established how geodemographic patterns are manifested in the living environments in which residents live and interact.

Street View Imagery (SVI) offers a promising complementary data source for capturing the visual dimension of the living environment. SVI has emerged as a means to gather detailed information on street layouts, building facades, and green spaces at a finer spatial scale (Biljecki and Ito, 2021; Kang et al., 2020). Leveraging this capability, existing studies have shown that SVI can effectively predict a range of single socio-demographic variables, such as income (Gebru et al., 2017; Suel et al., 2019), poverty (Wu et al., 2025), and health outcomes (Fan et al., 2023). However, these studies mainly focus on isolated vari-

*Corresponding author

ables, whereas geodemographic classification concerns multi-dimensional neighbourhood patterns jointly shaped by multiple demographic and socioeconomic characteristics (Lin, 2024). Therefore, it remains unclear how visual features derived from SVI relate to geodemographic patterns across urban areas.

This challenge is further amplified in cross-city analysis. Differences in historical contexts and cultural backgrounds may cause 'visual-demographic' association patterns to vary across cities. The same socio-demographic group may correspond to distinct visual environments, indicating cross-city heterogeneity in visual-demographic relationships. Therefore, models trained in one city exhibit reduced performance when applied to other contexts (Hou et al., 2025). These differences likely stem from underlying variations in visual features, which are essential for comprehending cross-city heterogeneity in visual-demographic relationships and for establishing a more generalizable framework.

However, very few studies have yet tested, across multiple cities, how far neighbourhood-level socio-demographic groups can be understood by Street-View-derived measures of the built environment, and scalable national SVI-based geodemographic models have yet to be established. These limitations reveal a critical gap in systematically integrating image-derived variables into geodemographic classification and evaluating its effectiveness and scalability. To systematically address this gap, this study aims to answer three research questions (RQ):

- RQ 1** To what extent can urban visual features be associated with socio-demographic group patterns within a city?
- RQ 2** Which visual features are most informative for distinguishing socio-demographic groups?
- RQ 3** To what extent are visual-demographic relationships transferable across cities with different urban contexts?

To address these questions, this study proposes a novel framework for understanding socio-demographic groups from urban visual data, such as SVI, and examining the effectiveness of establishing a nation-wide geodemographic classification model based on image-derived variables. The study focuses on 10 cities in England and Wales. First, we extract five categories of image-derived variables from SVI: semantic categories, object categories, scene categories, pixel categories, and spatial categories. Second, this study employs multiple machine learning models to train an optimal classification model using UK Output Area Classification (OAC) data, an official census-based geodemographic data that assigns each output area to a socio-demographic group based on socio-demographic variables. Finally, we conduct cross-city model transfers to further evaluate the model's robustness and generalizability and to examine variations in visual-demographic relationships among cities. This study aims to contribute to the development of a visual data-driven framework for geodemographic classification in social geography, and provides a

finer-grained understanding of the neighbourhood environments associated with different socio-demographic groups, thereby helping to identify the specific environmental needs of local communities and supporting more targeted and equitable place-based planning and policy interventions.

2. Background and Related Work

2.1. Traditional data sources for geodemographic classification

Geodemographic classification studies employ diverse data sources to characterize differences among socio-demographic groups. Each data source offers distinct advantages not only in representing socio-demographic attributes but also in functional characteristics and activity patterns.

Traditional geodemographic classification studies primarily rely on census data, such as the UK's Output Area Classification (OAC) in 2001, 2011, and 2021 (Vickers and Rees, 2007; Gale et al., 2016; Wyszomierski et al., 2024). Due to its broad coverage and interpretability, census data form the foundation for geodemographic classification. Although census-based geodemographic classifications reliably represent the overall population structures, they are primarily based on aggregated social statistical variables. To enhance the representation of this information, recent research has incorporated additional data sources, such as remote sensing imagery, to supplement the characterization of regional spatial features (Luo et al., 2022; Batista e Silva et al., 2020). However, remote sensing data is often insufficient to reveal the activity patterns and detailed social processes within communities. Consequently, studies have incorporated mobility data to provide additional information on population movement (Mavrogeni et al., 2025; Gao et al., 2021; Uğurel et al., 2026; Mavrogeni et al., 2026). Recent bespoke geodemographic classifications have further demonstrated how these emerging data sources can be incorporated to characterize specific dimensions of urban populations and activities, including neighbourhood interactivity, night-working patterns, and financial precarity (Ye and Singleton, 2026; Mavrogeni et al., 2025, 2026). Building on these data sources, recent research has explored participatory data from social media platforms, including Twitter and Facebook, to further reflect geodemographic classification (Chappell et al., 2017; Gulma, 2022). Social media data offers greater real-time insights than traditional census data and also reflects social interactions. For example, Gulma (2022) combined social media data with census variables to propose a novel geodemographic classification framework. However, social media data is subject to platform participation bias and is highly context-dependent, which complicates the reliable characterization of socio-demographic group environments.

As data sources in geodemographic research become increasingly diverse, there is a corresponding demand for methods capable of effectively representing and integrating heterogeneous and complex information. Machine learning techniques have recently been adopted in geodemographic research to support the representation and analysis of complex data relationships. For example, De Sabbata and Liu (2023) utilized a graph neural

network to generate geodemographic representations that account for spatial relationships. Singleton and Spielman (2026) further emphasized the potential of learned representations to capture complex nonlinear relationships and integrate diverse data types.

Within this context, SVI presents a promising data source for geodemographic research, as its high-dimensional and unstructured visual information can be represented and analysed using machine learning approaches (Suel et al., 2019). In addition, SVI records the visual environment at street level and provides diverse visual information from a human perspective (Biljecki and Ito, 2021), which is often inaccessible through census data, remote sensing, and other multi-source data. Therefore, SVI provides valuable supplementary data to support a more comprehensive understanding of geodemographic classifications.

2.2. Street view imagery as a potential source for geodemographic classification

Existing SVI-based studies related to geodemographic classification can be broadly divided into two categories, with most of them focusing on socio-demographic variables that constitute of geodemographic classification. The first category involves directly inputting raw SVIs into deep learning models to predict socio-demographic variables, such as income, education, unemployment, housing, health, and crime. For example, Suel et al. (2019) utilized raw images to predict various socio-demographic variables without extracting intermediate defined features. Studies indicate that the model has better performance in inferring the quality of the living environment and mean income using large-scale SVIs. However, although SVI-based prediction methods are effective for socio-demographic variables inference, the direct learning of representations from images often overlooks which specific visual features are most influential.

The second category first extracts semantic or object-level features from SVIs and then uses these features for socio-demographic variables inference. This method typically uses computer vision methods such as semantic segmentation and object detection. For example, features including greenery, sky, vehicles, and typefaces can be extracted from SVIs, and the relationships between these features and variables such as income and socioeconomic status (Li et al., 2015; O'Regan et al., 2021; Gebru et al., 2017; Ma et al., 2019), health, and walkability can be analysed (Wang et al., 2019; Yin and Wang, 2016). This approach offers greater interpretability, as results can be mapped to specific visual features. For example, Li et al. (2015) demonstrated that the level of greenery differs among residents of various social classes, with low-income groups residing in areas with reduced greenery. While existing research has established that SVIs can infer multiple socio-demographic variables, their application in geodemographic pattern inference remains limited. Moreover, many studies focus on a small number of semantic objects or single environmental indicators, resulting in a localized portrayal of socio-demographic differences in street views. Places are generally perceived through the integration of multiple visual categories rather than any single feature (Zhang et al., 2021). These categories may include semantic composition, object distribution, scene attributes, and other multi-

dimensional visual features. Therefore, integrating multi-dimensional visual information into geodemographic classification research enables a more comprehensive depiction of the visual environment and enhances the understanding of geodemographic classifications.

3. Methodology

The developed framework consists of three steps (Figure 1):

1. Extraction of image-derived variables from SVIs. This step extracts image-derived variables in five categories based on Google Street View (GSV) and computer vision: semantic category, object category, scene category, pixel category, and spatial category.
2. Constructing the prediction model for socio-demographic groups. Multiple machine learning models are trained based on the image-derived variables in SVIs and the UK census units.
3. Evaluating cross-city transfer bias. This step applies the trained model for each city to other cities for transfer, thereby observing the similarity and transferability of each city in the socio-demographic group inference task.

3.1. Study area and data

This study focuses on 10 cities in the United Kingdom: Birmingham, Bristol, Cardiff, Leeds, Liverpool, London, Manchester, Nottingham, Sheffield, and Southampton. These cities have been chosen for their varied historical development paths, diverse socio-economic roles, and distinct urban forms, which are essential for understanding nation-wide urban socio-demographics, rather than being constrained by a particular city.

As core national and regional cities, London, Manchester, and Birmingham have multifaceted roles, and intricate socio-demographic structures typical of major urban centres (Sassen, 2013; Barber and Hall, 2008; Peck and Ward, 2002). Sheffield, Leeds, and Nottingham, with their coexistence of traditional industrial and university town characteristics, represent mid-sized cities undergoing industrial transformation and renewal (Dabinett, 2005; Meegan, 2015; Crewe and Beaverstock, 1998). Liverpool and Southampton, as key maritime cities, have had their growth significantly shaped by international commerce and migration, leading to notable demographic diversity and spatial disparities (Sykes et al., 2013; Pinch, 2002). In contrast, Bristol, a pivotal hub in southwest England, retains its historic port and manufacturing layout due to its industrial past, while evolving to include a varied socioeconomic fabric (Tallon, 2007). Meanwhile, Cardiff, as Wales' capital, reflects typical demographic and spatial characteristics influenced by its roles in administration, politics, and culture (Waite, 2015). By encompassing a diverse range of cities, this study aims to test the model's applicability across multiple urban contexts and systematically analyse the socio-demographic heterogeneity revealed by cross-city transfer biases.

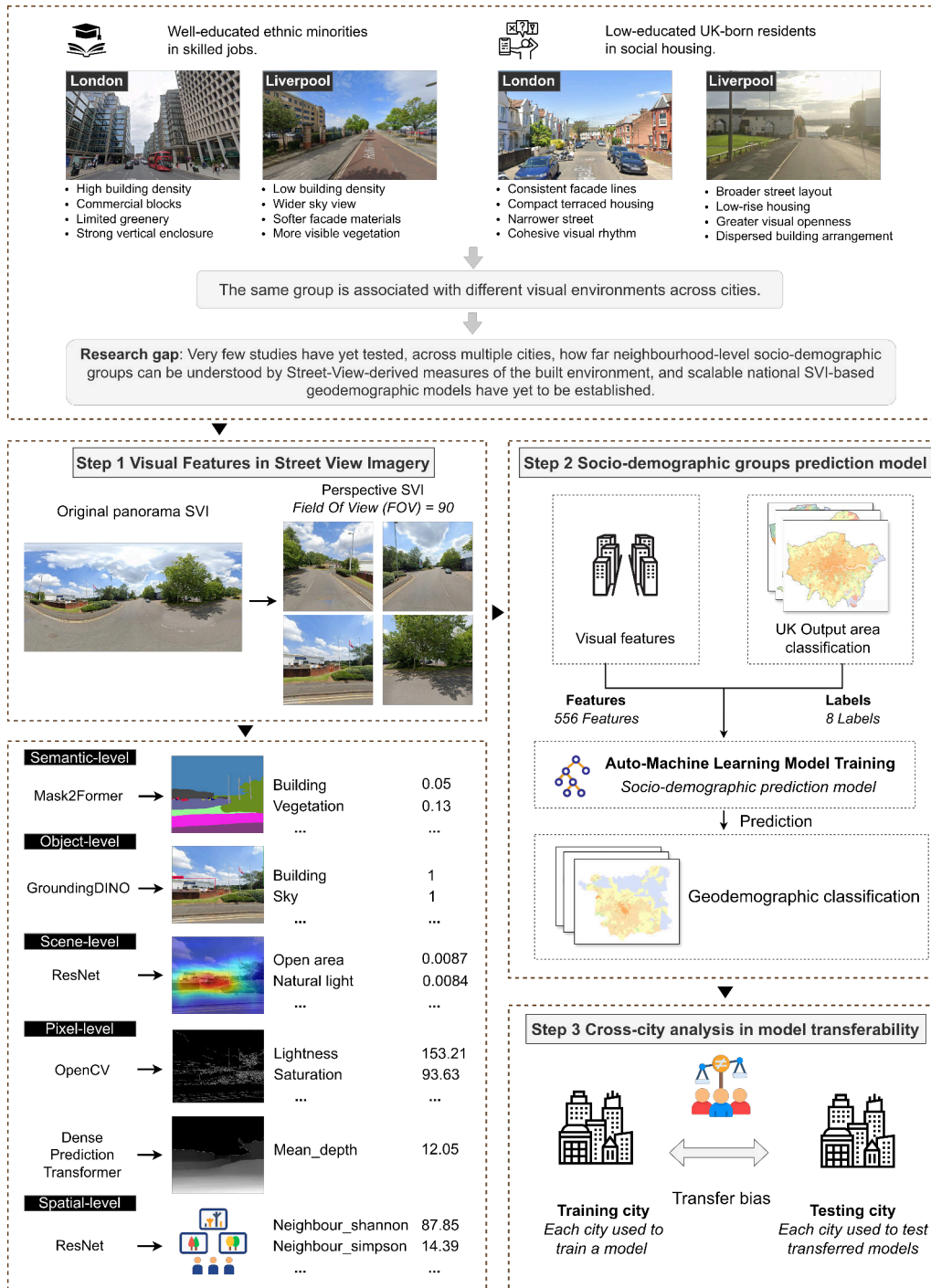


Figure 1: The methodology we introduce to infer socio-demographic information and evaluate it: from image-derived variables extraction from street view imagery (step 1) to socio-demographic group model training (step 2) to cross-city analysis (step 3).

The area classification data is based on the UK Official Output Area Classification (OAC), a national geodemographic classification system developed by the country's Office for National Statistics (ONS) (Wyszomierski et al., 2024). Relying on comprehensive census data, the OAC integrates 60 core variables using k -means clustering analysis, covering a wide range of dimensions such as demographics (gender and age), ethnicity and origins (nationality, race, birthplace, language), employment (occupation, profession, trade or employment), usual residence (place of abode, character of dwelling) and religious beliefs, among other dimensions, to construct an urban area classification system. The classification system categorizes the OAC units into eight socio-demographic groups: Retired Professionals (Group 1), Suburbanites and Peri-Urbanites (Group 2), Multicultural and Educated Urbanites (Group 3), Low-Skilled Migrant and Student Communities (Group 4), Ethnically Diverse Suburban Professionals (Group 5), Baseline UK (Group 6), Semi- and Un-Skilled Workforce (Group 7), and Legacy Communities (Group 8). Figure 2 illustrates the clustering variables and the dataset across the study area, together with an interpretation of each group.

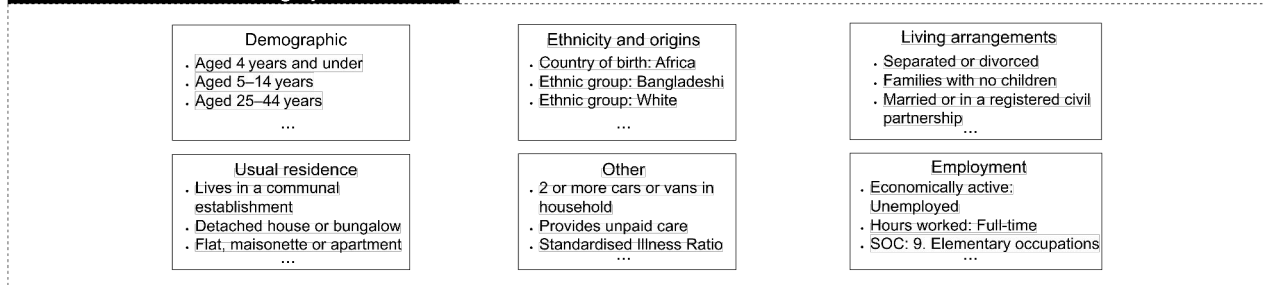
The OAC classification system has been widely validated. Existing studies have demonstrated the OAC classification system's effectiveness in studying ageing population (Yang et al., 2023), health (Petersen et al., 2011), and public safety (Anderson, 2010; Gulma, 2022) across diverse urban contexts. In this study, the OAC system provides a basis for an objective and standardized demographic classification. Significant differences in demographic characteristics among different groups provide an analytical basis for examining the visual-demographic relationship in different urban contexts, which also strengthens the methodological validity.

The approach relies on Google Street View (GSV) as a source of SVI to extract image-derived variables. SVI is characterized by high spatial coverage and high image quality, which can provide more accurate information about the urban visual environment (Biljecki and Ito, 2021). In this study, we first generated sampling points using a systematic spatial sampling method based on the H3 hierarchical hexagonal spatial indexing system, ensuring a regular and relatively even spatial distribution across the study areas. We adopted H3 resolution 10, corresponding to an average cell area of 0.015 km².

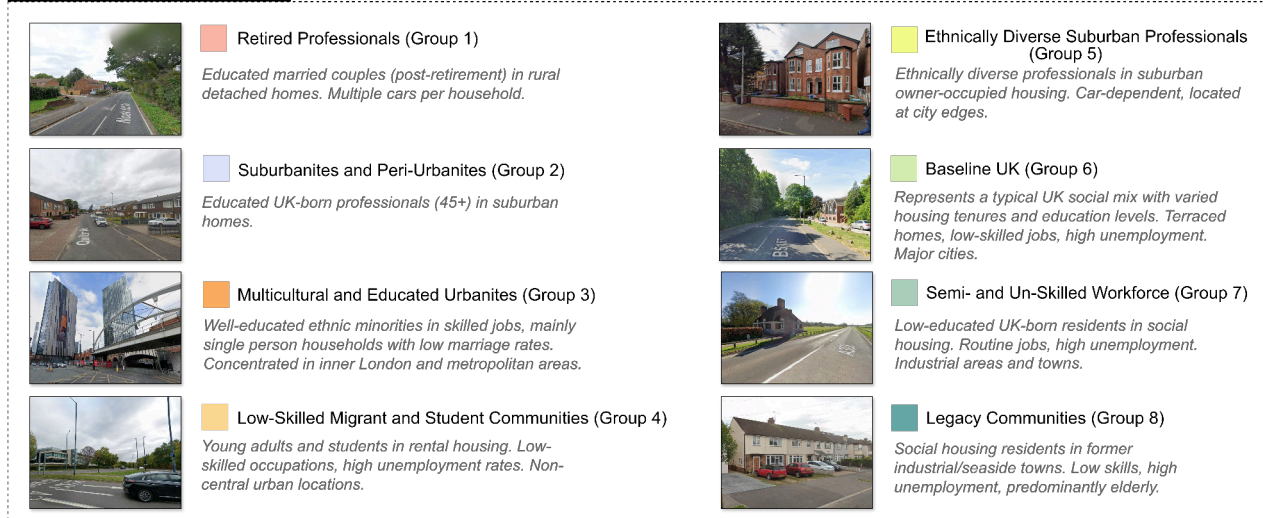
For each sampling point, we first retrieved SVI captured in 2024, while images captured between 2021 and 2023 were treated as candidate images to improve the spatial coverage of SVI. Finally, we obtained 636,086 SVIs collected in 10 UK cities between 2021 and 2024.

Each image was subsequently assigned to the corresponding OA according to its actual capture location using a point-in-polygon spatial join. Images located close to an OA boundary but whose capture coordinates remained within a particular OA were assigned to the OA containing the capture point. Because OAs vary in area and shape, the number of street-view samples within each OA also differed. To illustrate this variation, we further plotted the distribution of SVI sampling density (the number of SVI samples per unit OA area) for each city (Please refer to Appendix B).

Census-derived socio-demographic variables



Socio-demographic groups



Study area

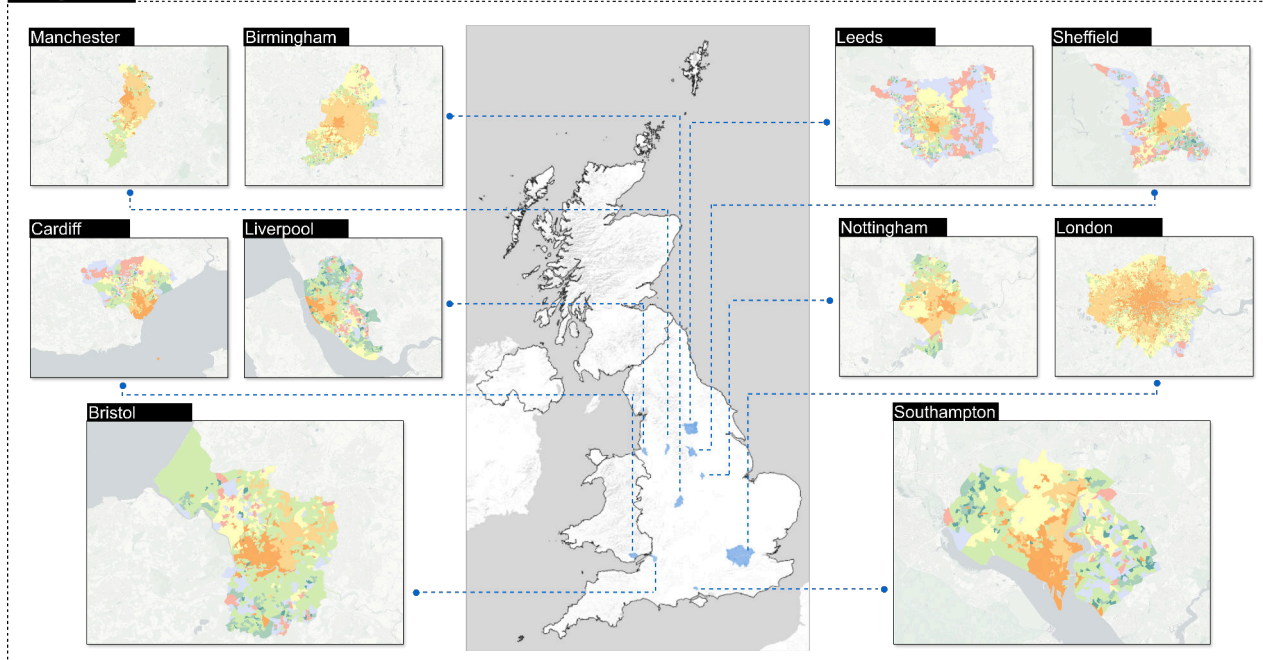


Figure 2: Study area and the clustering variables on demographic groups that we are predicting in this work with street view imagery. We focus on 10 diverse cities in England and Wales.

3.2. Visual features in street view imagery

SVI comprehensively reflects the visual features of the built environment from a pedestrian's perspective and it is an important data source for urban research (Biljecki and Ito, 2021; Zhang et al., 2023). Perspective views are created by cropping each GSV panorama (Figure 1) to reflect the human visual range. Therefore, first, a perspective transformation on panoramic SVI is performed, setting the field of view (FOV) to 90 degrees. Furthermore, this study utilizes ZenSVI (Ito et al., 2025), an open-source Python package, to systematically extract multi-dimensional visual features from SVI, ultimately generating 556 image-derived variables (Figure 1).

As shown in Table 1, image-derived variables in five categories are extracted: semantic, object, scene, pixel, and spatial categories. For the semantic category, we use the Mask2Former model, pretrained on the Cityscapes dataset (Cheng et al., 2022a), for semantic segmentation of each image and obtain the relative proportion of each semantic category in the image, such as road, vegetation, and sky. Then, the mean and standard deviation of these proportions are aggregated to the OAC parcels. For the object category, we use GroundingDINO pretrained on COCO, O365, and Open Images (Liu et al., 2024) for object detection. A total of 32 object categories are defined for detection, including bicycles, cars, and buildings, and the detected objects are aggregated by their total occurrence counts within each unit. For scene attribute extraction, this study uses a ResNet model pretrained on the Places365 dataset to calculate the probabilities of 365 scene categories (such as residential areas, building facades, and field roads) and the scores of 102 scene attributes (including features such as bicycle areas, asphalt pavement, and grass cover) (Zhou et al., 2017). These outputs are then jointly normalized using a softmax function and averaged across all images within each unit to serve as scene features. Six scene diversity indicators (including the Shannon and Simpson diversity indices, etc.) are calculated for each unit from these averaged values of all scene categories and attributes (Nagendra, 2002). For the extraction of pixel category, we calculate 9 types of features, including hue, saturation, depth, etc. (Ranftl et al., 2021; Ito et al., 2025), and average them across all images within each unit. Finally, for the spatial category, this study calculates 4 types of spatial features (including the mean area-weighted Shannon and Simpson diversity of adjacent parcels, etc.). For the complete list of features, please refer to Appendix A.

We selected the five complementary information dimensions described above to jointly construct a visual representation of place. These dimensions characterize the image composition of a place through semantic segmentation, the number of visible objects through object detection, the overall environment and scene attributes through scene-level features, visual appearance such as image colour and texture through pixel-level features, and the surrounding spatial environment and scene diversity through spatial features. These dimensions characterize the street environment at different levels and together form a multi-dimensional visual representation of the visual environment. Existing studies have shown that visual elements and their composition in street environments, visible objects, overall scenes, and visual characteristics such as colour and texture are associated with people's

perceptions and evaluations of places (Zhang et al., 2018b; Verma et al., 2020; Rossetti et al., 2019; Zhao et al., 2023), while the spatial context and visual heterogeneity among neighbouring streetscapes can also influence the overall perception of the local environment (Yang et al., 2026). Therefore, integrating these complementary visual dimensions can provide a more comprehensive representation of the visible environment of a place.

Table 1: Summary of image-derived variables by each category.

Visual feature categories	Image-derived variables summary
Semantic	38 semantic variables: • Road, vegetation, sky, etc.
Object	32 object variables: • Bicycle, car, building, person, etc.
Scene	473 scene-related variables: • 365 scene categories • 102 scene attributes • 6 scene diversity indices
Pixel	9 low-level variables: • Hue, saturation, lightness, mean depth, etc.
Spatial	4 neighbourhood diversity variables: • The mean area-weighted Shannon and Simpson diversity of adjacent parcels • The mean Shannon and Simpson diversity per square kilometre of adjacent parcels

3.3. Predicting socio-demographic groups using multi-dimensional image-derived variables

The OAC socio-demographic labels from Wyszomierski et al. (2024) provide the target variables for this classification task, while image-derived variables obtained from SVI and aggregated at the 2021 OAC units (Section 3.2) serve as the input for the prediction model.

This study employs the AutoGluon framework (Erickson et al., 2020) for model training. The selected variables are high-dimensional and exhibit strong non-linearity, with pronounced distributional differences across cities. As a result, manual hyperparameter tuning is often inadequate to achieve robust optimization. AutoGluon, an automated machine learning (AutoML) toolkit for tabular data, systematically compares multiple candidate learners such as tree-based models and neural networks within a unified training

pipeline. It integrates their complementary strengths through multi-layer stacking and applies automated hyperparameter optimization, including Bayesian search, to tune key parameters and enhance generalization. This approach reduces subjectivity in model selection, enabling more stable model performance with minimal human intervention.

Specifically, we train a separate model for each city and evaluate multiple classification algorithms to assess the predictive performance of image-derived variables across diverse urban contexts. During model training, the data is divided into a training subset and a test subset in an 80%:20% split. To further examine the generalizability of visual features for predicting socio-demographic group distributions at a broader spatial scale, we also combine data from 10 cities to construct a national model and test it with different ratios of testing set and training set. This approach aims to characterize the model’s generalization performance and to determine whether, with limited training data, the model can learn stable visual–demographic associations that enable effective prediction in previously unseen areas. Such capability may offer a reference framework for visual-driven geodemographic classification in areas lacking sufficient official socio-demographic statistics.

3.4. Cross-city analysis in model transferability

To further reveal model stability, generalization, and the heterogeneity of visual–demographic relationships across cities, we conduct cross-city model transfer. Each city’s model is not only validated on a local test set, but the model is also applied to datasets from nine other cities without re-training or fine-tuning. This approach allows us to quantify the extent to which visual–demographic relationships learned in one city generalize to others. However, model performance degrades when transferred to a different context (Hou et al., 2025). This issue is commonly referred to as transfer bias, a general phenomenon in machine learning.

In this study, for each current city (i.e. testing city) i , transfer bias is defined as the difference between the accuracy of the local model trained and tested in city i and the accuracy of a model trained in another city j and tested in the city i , as shown in the formula below:

$$TransferBias_{ji} = A_{i,i} - A_{j,i}$$

where $A_{i,i}$ represents the accuracy of the local model for city i , and $A_{j,i}$ represents the accuracy obtained when a model trained in another city j is applied to city i . This metric quantifies the extent to which the model’s effectiveness declines in outlying cities relative to its local performance.

A smaller transfer bias indicates greater similarity in the visual-demographic relationships between the two cities and higher transferability. In contrast, a larger transfer bias suggests more differences in their visual–demographic relationships. Therefore, transfer bias provides new quantitative evidence for understanding the heterogeneity of the visual–demographic relationship across cities, helping to develop a more generalizable framework.

4. Results

4.1. Model performance

4.1.1. Performance of local models

To develop more effective machine learning prediction models, we compared a series of machine learning models and trained them using AutoGluon. L1 represents the initial training results of each basic algorithm (CatBoost, Extra Trees, LightGBM, NeuralNetFastAI, NeuralNetTorch, Random Forest, and XGBoost). L2 represents the training results obtained through further training and parameter updates based on the L1 model, aiming to improve the model’s generalization performance and stability.

Table 2 shows the model evaluation accuracy of various machine learning models on L1 and L2 for the 10 cities selected for this study (Weighted-F1 results are provided in [Appendix C.1](#)). During model training, we used an 80%:20% ratio of training to test data. LightGBM performed well in most cities. Furthermore, we observed that L2 with further parameter updates outperformed L1 in most cities, suggesting that further parameter optimization can improve the model’s generalizability for complex visual features. Model performance varied across cities, with London achieving an overall accuracy of 0.7031, while Birmingham, Manchester, and Nottingham achieved accuracies above 0.6000. In addition, when averaged across all 10 cities, the LightGBM model achieved an accuracy of 0.5679. This accuracy result demonstrates that models based on image-derived variables can effectively infer socio-demographic groups.

Table 2: Comparison of model performance (accuracy) across the 10 cities using different machine learning algorithms under two training stages. Bold numbers indicate the highest accuracy achieved for each city across all models.

Model	Training stage	Birmingham	Bristol	Cardiff	Leeds	Liverpool	London	Manchester	Nottingham	Sheffield	Southampton
CatBoost	L1	0.6074	0.5657	0.5092	0.4707	0.5097	0.6890	0.6013	0.5838	0.4294	0.4834
	L2	0.6352	0.5839	0.4862	0.4922	0.5129	0.6982	0.6144	0.6091	0.4654	0.4901
Extra Trees	L1	0.5672	0.5438	0.4541	0.3750	0.4516	0.6575	0.5882	0.5888	0.3934	0.4768
	L2	0.6291	0.5730	0.4862	0.4805	0.4839	0.6971	0.6144	0.5990	0.4488	0.4901
LightGBM	L1	0.6229	0.5657	0.5138	0.4531	0.5065	0.6945	0.6144	0.5990	0.4321	0.5033
	L2	0.6321	0.5766	0.5229	0.4766	0.5258	0.7031	0.6340	0.6142	0.4903	0.5033
NeuralNetFastAI	L1	0.6198	0.5766	0.4633	0.4746	0.4774	0.6962	0.6275	0.6091	0.4432	0.5099
	L2	0.6198	0.5730	0.4450	0.4648	0.4742	0.6967	0.6307	0.6091	0.4571	0.5364
NeuralNetTorch	L1	0.6182	0.5839	0.5229	0.4844	0.5065	0.6935	0.6242	0.6447	0.4737	0.5762
	L2	0.6275	0.5766	0.5046	0.4727	0.5065	0.7020	0.6340	0.6497	0.4820	0.5430
Random Forest	L1	0.5920	0.5401	0.4679	0.4336	0.4968	0.6746	0.5850	0.5635	0.4211	0.4901
	L2	0.6306	0.5839	0.5000	0.4941	0.5258	0.6979	0.6209	0.5990	0.4848	0.5033
XGBoost	L1	0.6151	0.5547	0.4954	0.4375	0.4806	0.6907	0.5850	0.6041	0.4183	0.4834
	L2	0.6260	0.5657	0.4862	0.4766	0.5032	0.6992	0.6046	0.6041	0.4820	0.5099

To further illustrate the spatial distribution of prediction errors, we visualize prediction correctness across cities using binary maps. Figure 3 shows the prediction results when

we apply each local model to its own full dataset (including training samples) to obtain spatially complete maps, and the precision for each socio-demographic group. The results indicate that, in most cities, misclassifications are predominantly sparsely distributed. We note that the London model achieves the highest predictive accuracy on the test set (Table 2). However, the binary map displays a larger number of erroneous predictions. This may be attributable to substantial differences in overall dataset across cities and within-city sample sizes (Please refer to Appendix D), making errors appear more numerous in London due to denser samples. In addition, some cities contain extremely small sample groups. These minority groups provide limited information for effective learning and are therefore more prone to misclassification in visual environments. Moreover, even when evaluation is conducted within the same city, the definitions of certain socio-demographic groups are inherently subject to limitations. We further analyse group definitions using cosine similarity (Please refer to Appendix E). The results show relatively high similarity between some groups, suggesting that groups are not fully separable in terms of their conceptual boundaries and descriptive features. Therefore, even when deploying a model within its own city, confusion between groups can still occur, leading to misclassification within cities.

4.1.2. Performance of the national model

To further develop a more generalizable framework, this study fuses data from 10 cities and trains a national model to evaluate the generalizability and robustness of image-derived variables at a larger spatial scale. In addition, we also train models using varying proportions of the training set. As shown in Table 3, the accuracy of each model is around 0.60. These results demonstrate that when fusing data from different cities, the socio-demographic prediction model based on image-derived variables maintains high classification accuracy, further confirming that image-derived variables can effectively infer the distribution of socio-demographic groups (The performance of each socio-demographic group in the national model is shown in Appendix C.2.). Among all training set proportions, LightGBM achieves the highest accuracy at the 70% and 80% proportions, more than any other model, demonstrating that the gradient boosting structure has greater robustness and explanatory power for high-dimensional, nonlinear image-derived variables. It can balance feature differences in complex feature spaces, thereby learning visual-demographic relationships that are both generalizable and differentiated. Therefore, LightGBM is used for the subsequent experiments, including the ablation test, feature importance analysis, and transfer bias analysis.

As shown in Table 3, the benefit of increasing the training sample size across the models is marginal beyond approximately 30,000 samples (corresponding to the 80% training proportion used in the subsequent experiments). This suggests that, for future prediction tasks, a dataset of around 30,000 samples is sufficient to support robust and reliable model performance.

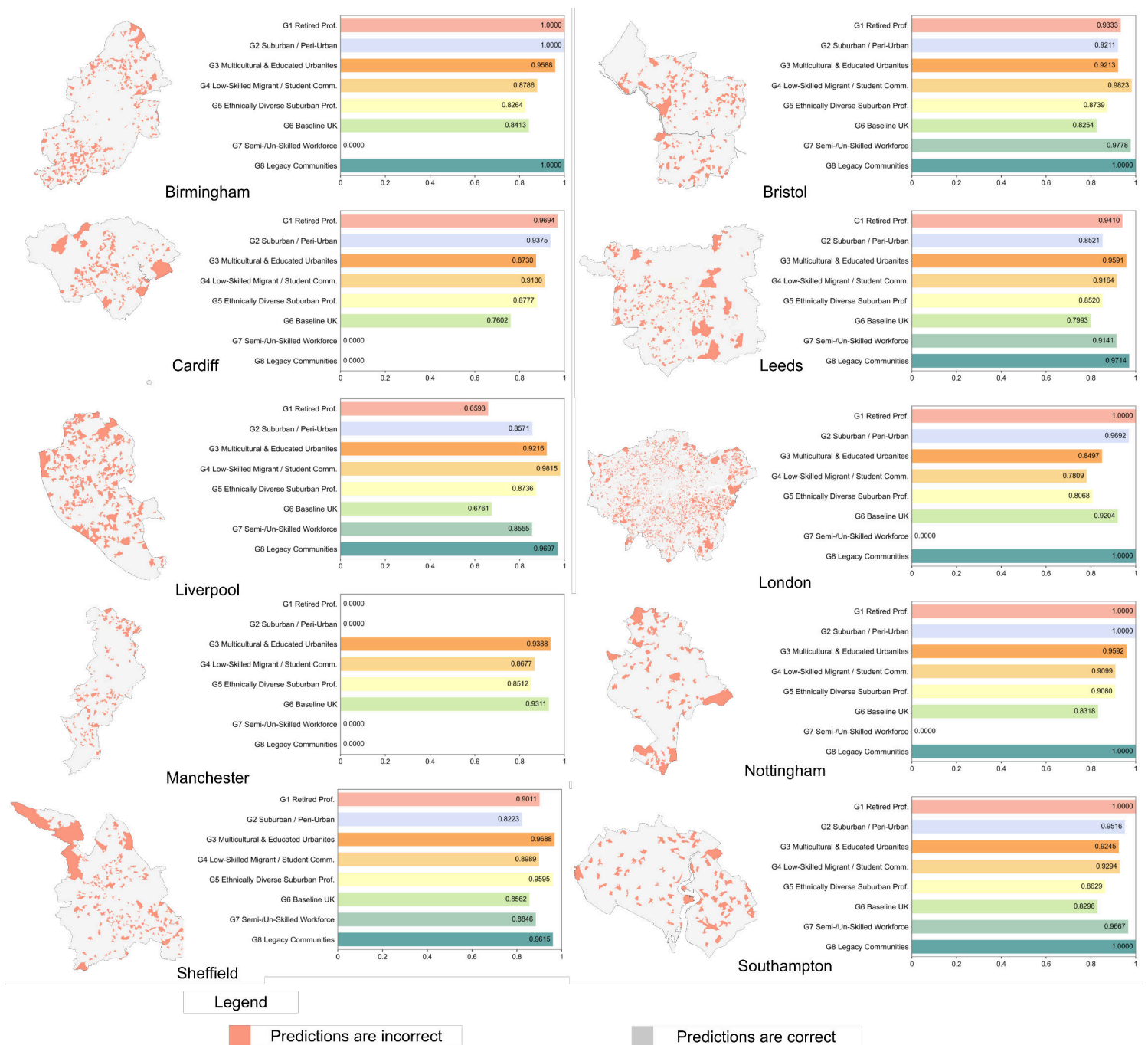


Figure 3: Within-city predictions of local models applied to their full datasets. We use binary maps to show the prediction correctness, and the precision for each socio-demographic group.

Table 3: Classification performance of models under different training data sizes. The training data is represented as the number of parcels, with the proportion of the full dataset ($N = 38,252$) shown in parentheses.

Model	Training stage	26,776 (70%)		28,689 (75%)		30,602 (80%)		32,515 (85%)		34,427 (90%)	
		Accuracy	Weighted-F1	Accuracy	Weighted-F1	Accuracy	Weighted-F1	Accuracy	Weighted-F1	Accuracy	Weighted-F1
CatBoost	L1	0.6020	0.5808	0.6029	0.5816	0.6033	0.5823	0.6086	0.5885	0.6064	0.5868
	L2	0.6172	0.6011	0.6157	0.6008	0.6149	0.5998	0.6182	0.5984	0.6179	0.6022
Extra Trees	L1	0.5523	0.5002	0.5516	0.4985	0.5571	0.5040	0.5605	0.5078	0.5549	0.5021
	L2	0.6101	0.5861	0.6114	0.5899	0.6139	0.5913	0.6190	0.5987	0.6134	0.5899
LightGBM	L1	0.6037	0.5812	0.6057	0.5834	0.6079	0.5872	0.6079	0.5880	0.6035	0.5832
	L2	0.6190	0.6028	0.6174	0.6023	0.6195	0.6035	0.6220	0.6077	0.6244	0.6097
NeuralNetFastAI	L1	0.6175	0.6015	0.6156	0.5975	0.6156	0.5999	0.6262	0.6116	0.6236	0.6057
	L2	0.6169	0.5983	0.6175	0.6022	0.6185	0.6034	0.6204	0.6070	0.6205	0.6027
NeuralNetTorch	L1	0.6135	0.5966	0.6112	0.5957	0.6114	0.5982	0.6173	0.6022	0.6218	0.6059
	L2	0.6172	0.6021	0.6177	0.6049	0.6164	0.6034	0.6220	0.6070	0.6208	0.6067
Random Forest	L1	0.5707	0.5275	0.5723	0.5294	0.5727	0.5279	0.5763	0.5332	0.5740	0.5324
	L2	0.6125	0.5946	0.6140	0.5958	0.6157	0.5981	0.6185	0.6031	0.6166	0.5996
XGBoost	L1	0.5979	0.5745	0.6155	0.6016	0.6033	0.5812	0.6114	0.5892	0.6009	0.5792
	L2	0.6178	0.6027	0.5990	0.5752	0.6176	0.6024	0.6228	0.6102	0.6257	0.6103

Figure 4 presents the alluvial plot of the national model to characterize its errors. It shows which groups the samples are predicted as. Because our study areas are relatively large cities in England and Wales, the distribution of OAC categories is uneven. Therefore, the prediction frequencies shown in the alluvial plot may be influenced by the underlying class distribution.

In addition to the effect of sample size, the observed prediction patterns may also depend on visual consistency within categories and visual distinguishability between categories. The same OAC category may contain heterogeneous built environments while still exhibiting certain distinctive visual characteristics that differentiate it from other categories. If a category overlaps substantially with other categories in terms of visual characteristics, it may remain difficult to identify even when it is relatively homogeneous internally.

We further select two groups with the highest prediction frequency (Groups 3 and 4) and show the corresponding SVIs. As illustrated in Figure 4, it can be observed that when samples from different groups are predicted as Group 3, these streetscapes tend to share key visual features that are highly consistent with the Group 3 (G3) ground truth (GT), leading to misclassification, such as higher building density and relatively narrower streets. Group 4 more frequently exhibits wider streets, with lower and more sparsely distributed buildings.

To more clearly illustrate the differences in visual environments across groups, Figure 5 displays the SVI across eight socio-demographic groups in different cities. The examples reveal that prediction patterns may arise from within-category consistency and limited distinguishability between categories. For example, Group 3 is frequently associated with relatively high-density built environments, providing recurring visual characteristics for classification. Ethnically Diverse Suburban Professionals (Group 5) are more

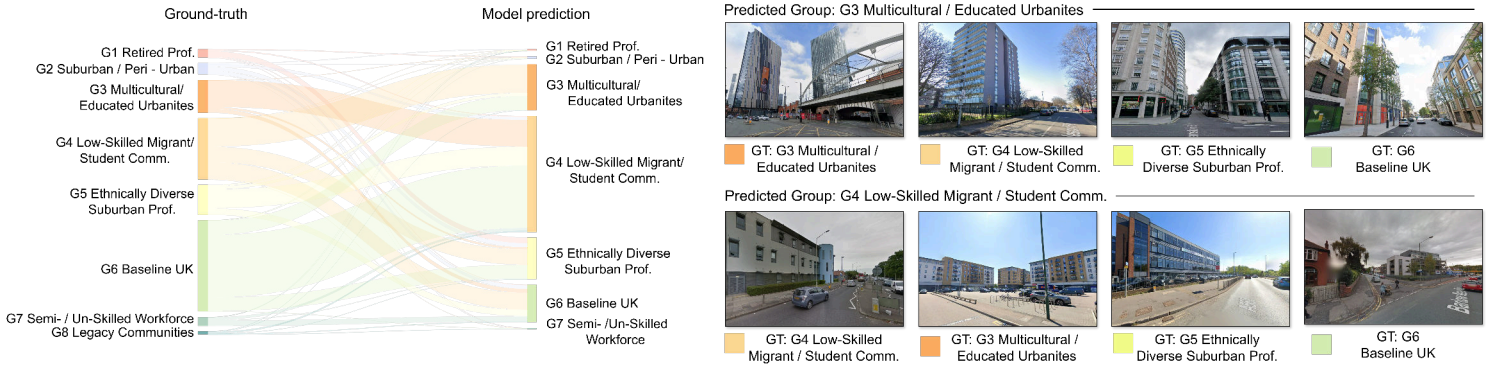


Figure 4: Alluvial plot of the national model with Ground Truth (GT) label and SVI examples for the two most frequently predicted groups. Source of imagery: Google Street View.

often associated with semi-detached houses with front gardens, often with cars parked outside. The visual heterogeneity across groups lays the foundation for the national model's predictions and its extension towards a nation-wide model.

Some categories share similar built environment characteristics, which can reduce their visual distinguishability. Although Legacy Communities (Group 8) are relatively homogeneous in socio-demographic terms, their low-rise housing, roads, and neighbourhood environments may visually overlap with other suburban or low-density categories. For example, both Group 8 and Ethnically Diverse Suburban Professionals (Group 5) can exhibit two- to three-storey residential environments. Visual heterogeneity between groups supports the stable performance of the national model during model training and its feasibility of scaling towards nation-wide inference, whereas visual similarity between groups determines a primary source of error for the national model.

4.2. Feature importance and interpretability

4.2.1. Global interpretability from ablation and importance

Figure 6 shows the results of an ablation experiment, which evaluates the impact of progressively removing visual features at different dimensions on model performance. The results show that, for a national model based on data from ten cities, progressively removing any feature dimension leads to performance degradation. Specifically, when the scene category is removed, accuracy drops to 0.5718, while when both the semantic category and scene category are removed, accuracy drops to 0.5465, a decrease of approximately 11.78%. Comparing the accuracy changes under the full feature category model and different ablation combinations demonstrates the effectiveness of this study's feature category selection.

To further understand the model's decision-making mechanism, this study outputs the feature importance value of each image-derived variable based on a LightGBM model on 10 cities' data. Panel A of Figure 7 shows the top 30 features and their importance values during model training. The results show that scene category accounts for 60%

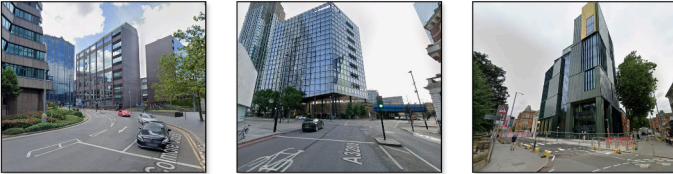
Retired Professionals (G1)



Suburbanites and Peri-Urbanites (G2)



Multicultural and Educated Urbanites (G3)



Low-Skilled Migrant and Student Communities (G4)



Ethnically Diverse Suburban Professionals (G5)



Baseline UK (G6)



Semi- and Un-Skilled Workforce (G7)



Legacy Communities (G8)



Figure 5: Visual examples of eight socio-demographic groups across different cities. Source of imagery: Google Street View.

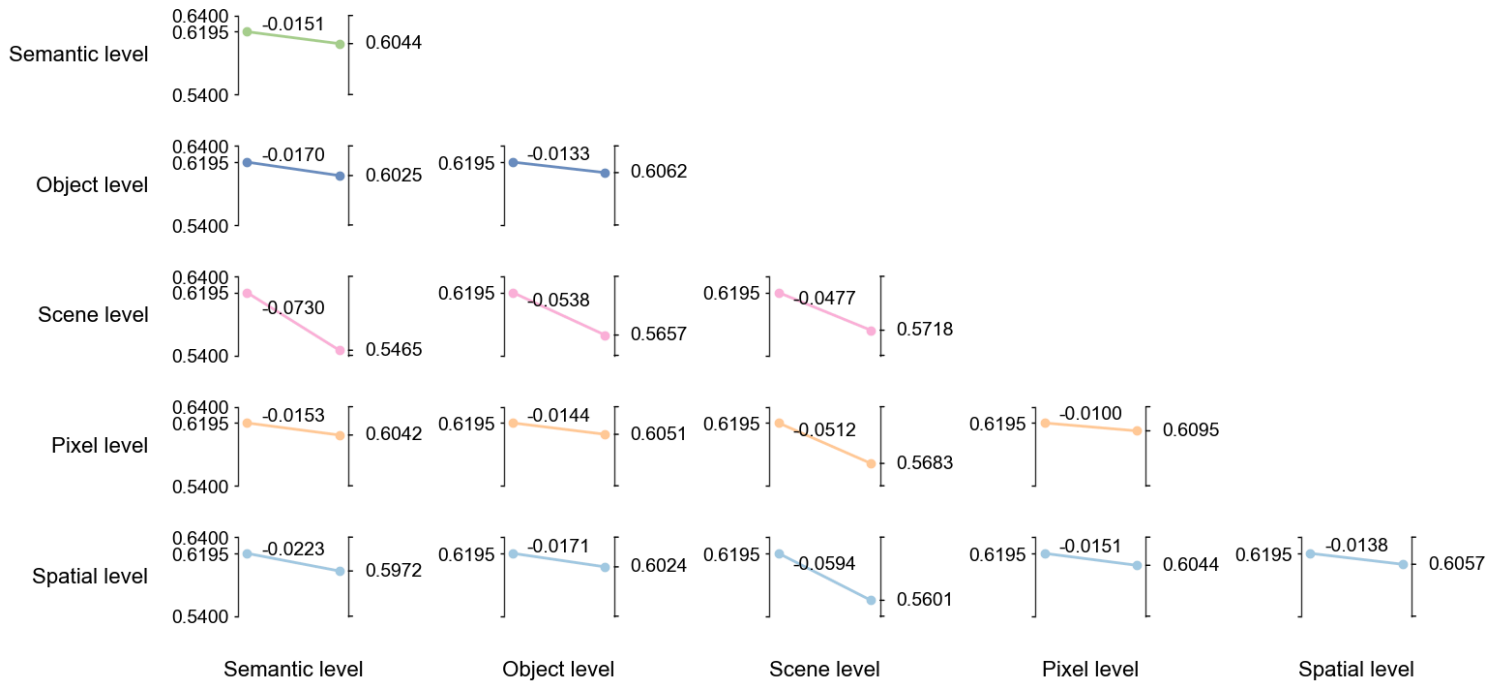


Figure 6: Ablation experiment results. Plots on the diagonal show the accuracy drop when deleting only one category. The horizontal axis represents the first category deleted, and the vertical axis represents the second category deleted. The numbers above the line represent the accuracy drop, while the numbers to the right represent the accuracy after ablation.

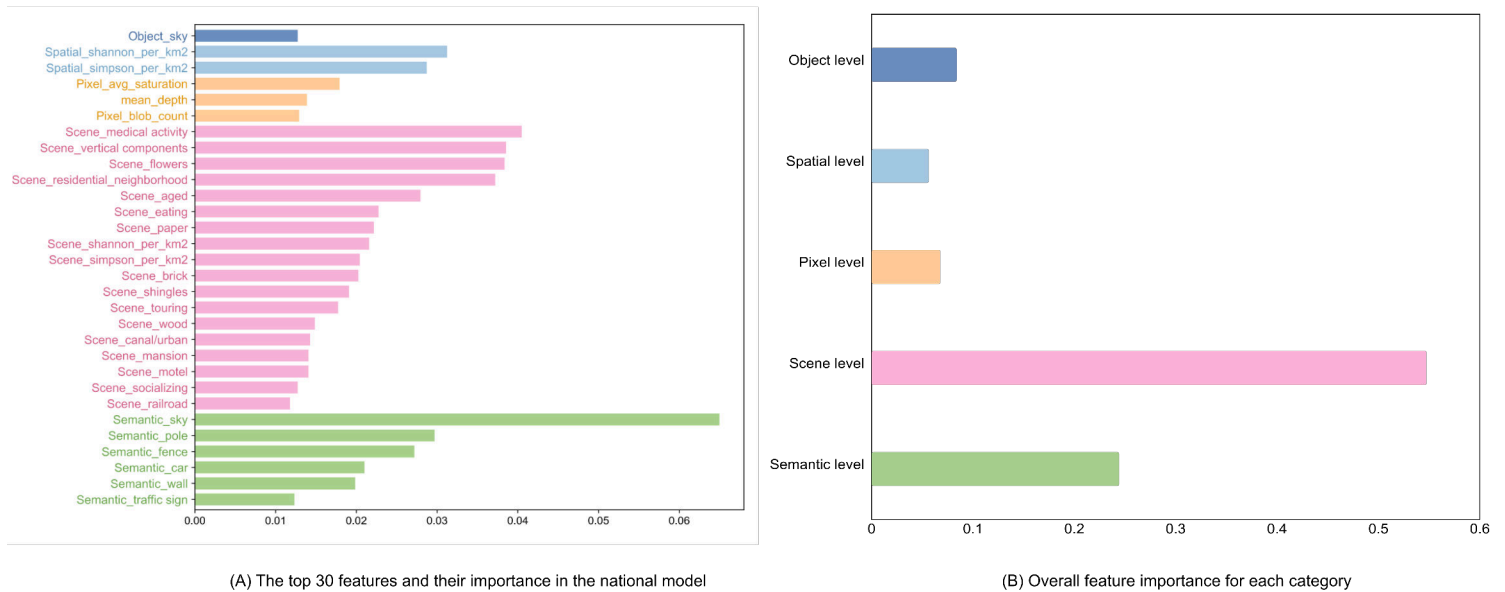


Figure 7: The feature importance in the national model.

of the top 30 features. Furthermore, this study computes and visualizes the total feature contribution of each category, as shown in panel B of Figure 7. The results indicate that the scene category contributes most significantly, followed by the semantic category. This observation is consistent with the ablation results, which show that removing the scene category causes the greatest performance degradation. Eliminating the semantic category also leads to a substantial decrease in model performance. The alignment between feature importance and ablation outcomes further supports the feature category selection in this study and demonstrates that urban landscape composition has explanatory power for the distribution of socio-demographic groups.

4.2.2. Group-level interpretability

The above results reveal the global contributions of visual features. However, different socio-demographic groups may display unique preferences for physical environments. We use SHAP to examine feature selection and attribution at the group level within the national model. Figure 8 displays, for each socio-demographic group, the top ten image-derived variables and their corresponding SHAP contributions, along with the aggregated contributions for each feature category.

For Retired Professionals (Group 1) and Suburbanites and Peri-Urbanites (Group 2), image-derived variables such as sky and spatial scene diversity contribute more to their classification. Grass also contributes to the classification of Group 1. These features capture differences in the residential environments of these groups. In contrast, although both Multicultural and Educated Urbanites (Group 3) and Low-Skilled Migrant and Student

Communities (Group 4) are defined by multicultural and migrant backgrounds, they are situated in distinct urban contexts. Scene heterogeneity and sky make relatively greater contributions to the classification of Group 3, whereas motel, flowers, and touring contribute more to Group 4. Ethnically Diverse Suburban Professionals (Group 5) show the largest contribution from residential scene features. Baseline UK (Group 6) represents the broad base of the United Kingdom’s social structure and is primarily characterized in the model by scene features such as parking garages, shingles, and brick, with the diversity of surrounding scenes also contributing. For Semi- and Un-Skilled Workforce (Group 7), features related to sky, grass, and smoke make relatively large contributions to the classification. For Legacy Communities (Group 8), grass, spatial scene diversity, and car features contribute relatively strongly.

4.3. Cross-city Model Transfer Performance

4.3.1. Transfer accuracy

To further assess the model’s stability and cross-city generalization, and to explore the transferability and similarity of visual features across diverse socio-spatial structures, this study transfers the local models to other cities without modification, revealing the heterogeneity of visual-demographic relationships between cities.

As shown in Figure 9, this study constructs a transfer matrix based on the transfer accuracy. The transfer bias defined in Section 3.4 can be obtained from panel A of Figure 9 as the difference between the diagonal and off-diagonal values, ranging from 0.03 (the Cardiff model applied to Southampton) to 0.48 (the Liverpool model applied to Birmingham). Furthermore, to quantitatively assess the transferability of local models, we calculate the average accuracy of each city model transferred to other cities, as well as the average prediction accuracy of each test city under different models. The results show that Leeds achieves the highest average accuracy when used as a training city and then transfers to the other nine cities, demonstrating strong generalization and transferability. Firstly, Leeds is located in the heart of England and may share similar characteristics of the built environment with other cities. Secondly, it is a city with a clear industrial history, having undergone a transformation from manufacturing to engineering and services, and in recent years, inner-city areas such as South Bank have also been undergoing regeneration and functional reshaping characterized by the introduction of commercial, residential, and public spaces (Meegan, 2015; Leeds City Council, 2022). This historical development path has resulted in an urban landscape containing streetscapes and neighbourhood types formed at different stages, which may enable the Leeds model to learn a richer visual-demographic relationship during training and facilitate the identification of similar patterns when transferred to other cities.

London, as the test city, achieves the highest average prediction accuracy, indicating that models from other cities perform well when transferred to London. According to its census, London is the most ethnically diverse region in England and has the highest proportion of residents born outside the UK (Office for National Statistics, 2022; Ethnicity

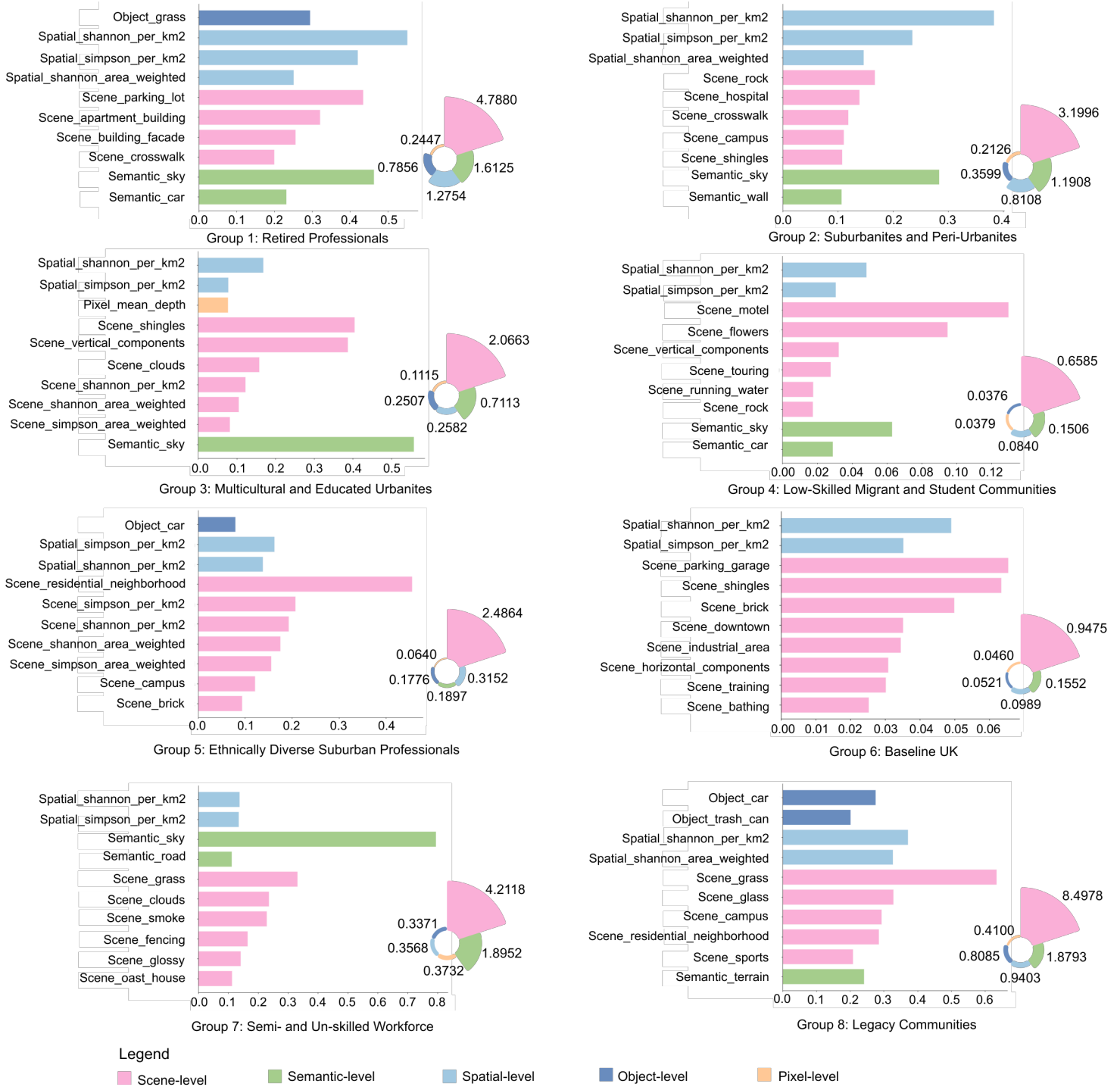


Figure 8: Feature contributions for each socio-demographic group, with the total feature contribution of each category.

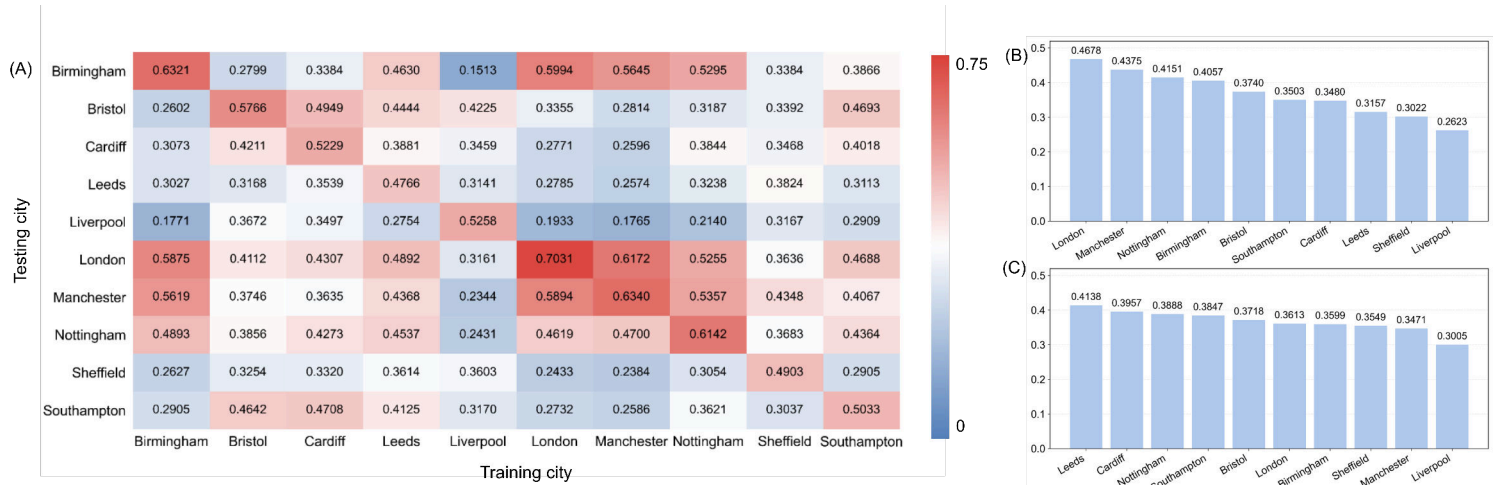


Figure 9: Transfer results for socio-demographic groups. (A) Transfer accuracy matrix. (B) Average accuracy when using the city as the test city. (C) Average accuracy when using the city as the training city.

facts and figures, 2022). This social context may be associated with a wider variety of neighbourhood types and street-level environments, making it easier for models trained in other cities to identify partially similar patterns when transferred to London.

In addition to the city-level transfer accuracy, we further provide confusion matrices to illustrate the prediction and misclassification patterns of each OAC category during cross-city transfer; please refer to [Appendix G](#).

4.3.2. Interpretation of transfer errors

To further understand the composition of transfer errors, Figure 10 shows the ground truth distribution of misclassified samples in cross-city transfer. Specifically, after transferring each local model to the other nine cities, we select all misclassified samples in each testing city and calculate the proportions of their ground truth labels. Notably, when Birmingham, London, and Manchester are used as testing cities, the misclassified samples predominantly come from Groups 3, 4, and 5, which include groups typically associated with higher levels of population mobility. The socioeconomic indicators (Please refer to [Appendix H](#)) suggest that these three cities exhibit higher migrant indicator and numbers of never-worked residents than the other cities. Existing studies have shown that continuous population inflows and outflows may alter the geodemographic structure of neighbourhoods and affect the conduct and interpretation of community-level surveys and analyses ([Hedman, 2011](#)). Residential turnover may also result in differences in geodemographic characteristics between newly arrived and long-term residents ([Hipp, 2020](#)). Meanwhile, SVI primarily records relatively stable built environments. Therefore, in areas experiencing rapid changes in population composition, the correspondence between current OAC labels and the visible environment may be weaker. Therefore, higher migration and mobility may increase the uncertainty of where these groups concentrate, making

these groups more difficult to identify for models trained in other cities.

In contrast, when Bristol, Cardiff, and Southampton are used as testing cities, the ground truth composition of misclassified samples is spread across a wider range of groups, yet Baseline UK (Group 6) accounts for the largest proportion for most training cities.

Furthermore, we visualize the spatial correctness of the transfer results (Figure F.15) to better understand the geographic distribution of misclassifications.

Specifically, London and Leeds have the highest average accuracies when the model is used as a testing city and a training city, respectively (London's local model also achieves the highest within-city accuracy), and we choose these two cities to analyse in-depth the predictions of transfer cases when applying these two cities' model to other cities (See Figure 11, Figure F.14 shows the complete transfer results between groups). Among them, when applying Leeds' model to Liverpool, the model predicts poorly. Taking a misclassified case as an example, an area in Liverpool belonging to Group 7 was incorrectly predicted as Group 6. By comparing the SVI of this misclassified area with a representative sample of Group 6 from Leeds, it is evident that the two exhibit high visual similarity in greenery and street layout, which likely misled the model. In addition, there are some differences between the two cities, such as labour force composition and economic activities (Table H.8), which might also affect the boundaries of the socio-demographic groups, making the effectiveness of the cross-city model application less effective. In addition, when applying London's model to Sheffield, the model misclassifies Sheffield's group 1 as group 4. From the physical environment, we can see that the two cities are relatively similar in openness and building types, but the two cities are very different in social indicators such as population density, demographic structure, and economic activity. However, a higher level of social and economic mobility is usually accompanied by different socio-economic groups, rates of urban regeneration, etc., which makes the visual-demographic correspondence more complex in London, and may lead to some misclassifications when applied to other cities

5. Discussion

5.1. A visual feature-based framework for understanding socio-demographic groups

This study demonstrates that multi-dimensional visual information contained in SVI can effectively support the understanding of socio-demographic groups. In contrast to traditional social-spatial analyses that rely on statistical data or functional indicators, SVI offers unique value because its visual content is directly linked to the physical characteristics of the collection location (Zhang et al., 2023). Contextual attributes, including variations in building facades, levels of greenery, and street width, are presented in concrete and directly observable forms within the street view, providing information on the visible street-level built environments associated with different socio-demographic groups. This aligns with the theory of human geography: place is a space that has been given meaning, and both its physical characteristics and perceptual qualities shape the lived world of

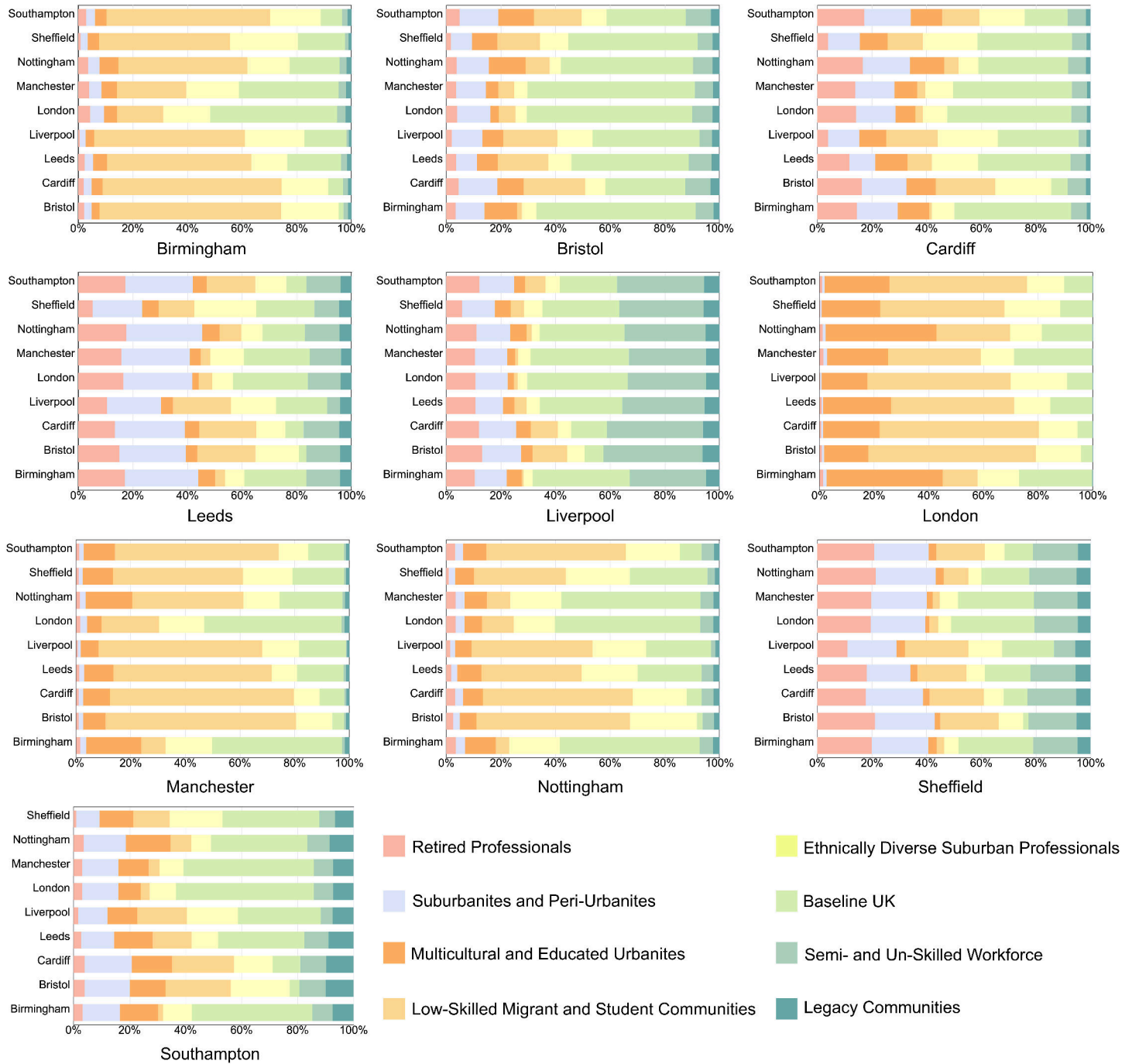


Figure 10: Composition of the ground truth for misclassified samples in cross-city transfer experiments. Each subplot represents a testing city evaluated using the local models of the other 9 cities (Training city). The bars indicate the percentage of each socio-demographic group within the total misclassified samples, highlighting the classes that are most prone to prediction errors during model transfer.



Figure 11: Case studies of model transfer in Leeds and London. The figure shows transfer cases with corresponding correctness maps and SVI examples. Source of imagery: Google Street View.

socio-demographic groups (Lipovac, 1997; Tuan, 1977). Despite their significance, such attributes have rarely been explicitly identified or examined in previous research.

Traditional socio-demographic data sources, such as censuses and sample surveys, may differ across geographical contexts in terms of variable availability, definitions, classifications, and geographical units, requiring additional harmonisation for cross-context applications (Coast et al., 2016). In this context, this study demonstrates the stability of the visual-demographic structure under varying training sample sizes, confirming the model's ability to maintain robust performance in highly heterogeneous environments. It also shows that, even in the absence of complete data, the understanding of socio-demographics can still achieve acceptable results using only partial training samples. This finding is particularly significant for future applications of visual-based demographic prediction in cities where ground-truth information is limited.

5.2. Heterogeneity in visual feature preferences across socio-demographic groups

The multiple dimensions of visual information within urban environments exhibit complementary relationships, collectively reflecting the visual preferences and environmental patterns associated with different socio-demographic groups. At the macro scale, scene descriptions and semantic compositions capture structural functions and landscape configurations (Zhang et al., 2018a). At the micro scale, pixel category and spatial category represent the textural characteristics of the built environment and the spatial continuity between parcels, respectively (Tang and Long, 2019). Serving as an integrative medium, SVI aggregates these diverse image-derived variables, enabling the model to establish connections across image-derived variables at different dimensions and thereby more comprehensively characterize the spatial distribution patterns of socio-demographic groups.

Socio-demographic statistics are primarily reported as population attributes, which limits the ability to directly examine interactions between population composition and the physical environment in shaping socio-spatial structures. However, the results demonstrate considerable heterogeneity in the associations between different socio-demographic groups and physical environmental contexts. This finding suggests that geodemographic classifications capture not only variations in population attributes but also distinct, observable streetscape characteristics identified through visual features. For instance, Multicultural and Educated Urbanites (Group 3) and Low-Skilled Migrant and Student Communities (Group 4) are associated with distinct urban contexts, with scene heterogeneity and sky contributing more to the classification of Group 3, while motel, flowers, and touring contribute more to Group 4. Despite the absence of explicit physical-environment variables in the official classifications, these differences further indicate that the model can identify environmental representations at the visual level that align with socio-demographic structures.

5.3. Transfer analysis as a new approach to urban heterogeneity

The cross-city transfer analysis offers a novel analytical perspective for understanding urban heterogeneity in visual-demographic relationships between cities. Liverpool

demonstrates less favourable performance in the experiments. Although geographically close to Manchester, Liverpool and Manchester do not display greater visual similarity or comparable visual–demographic structures. This finding indicates that visual resemblance between cities is not determined by geographical proximity but is more likely shaped by each city’s unique development trajectory, historical heritage, and planning institutions. [Pike et al. \(2024\)](#) pointed out that grouping different places under broad spatial labels may homogenize differences between places and obscure their distinct problems, development trajectories, and development potential. Therefore, the same OAC category may represent similar socio-demographic characteristics while being situated within substantially different local urban contexts, which may be reflected in differences in their built environments. For example, the development of the London-specific OAC indicates that national-scale classification systems may not fully capture the distinctive characteristics of particular urban contexts ([Singleton and Longley, 2024](#)). Therefore, cross-city transfer analysis also provides information on differences in local context underlying the socio-demographic similarities represented by conventional geodemographic classifications.

Additionally, misclassified areas may be associated with inter-city differences in migrant composition, labour structure, and economic activity levels. Recent geodemographic research has also begun to pay greater attention to urban activities and how cities function. For example, [Mavrogeni et al. \(2026\)](#) used dynamic mobility data to extend traditional population characterization based on places of residence and work to human interactions and urban functions, demonstrating that the spatial distribution of populations also needs to be understood in relation to activities and interactions within cities. In this study, transferring a model trained in London to Liverpool results in the lowest performance among all transfers of the London model. These cities differ significantly in migrant composition, and London also has a higher population density. Thus, even when visual appearances are similar, variations in socioeconomic factors can produce distinct spatial patterns of socio-demographic groups, which may reduce model transferability. While socioeconomic differences may partially explain model misclassification, these variables are not the sole determinants. Geodemographic classification patterns are influenced by multiple factors. Urban governance and policy interventions, for example, may also affect the relationship between visual features and socio-demographic attributes, thereby impacting the model’s capacity to interpret visual features during cross-city transfer.

5.4. Implications of SVI-based geodemographic classification

From a policy perspective, the UK’s National Planning Policy Framework (NPPF) ([Ministry of Housing, Communities and Local Government, 2024](#)) states that planning decisions address the social, economic, and environmental needs of populations as part of its sustainable development principles. Achieving this objective requires planners to understand not only the socioeconomic composition of populations, but also the built environment associated with different socio-demographic groups. However, traditional census and socioeconomic data do not adequately capture the built environment characteristics of

various socio-demographic groups. Our results address this gap by identifying the relationship between multidimensional visual characteristics and geodemographic classification, showing that different socio-demographic groups are associated with different visual characteristics. This indicates that SVI-based geodemographic classification can complement conventional data sources by providing finer-grained street-level evidence for more context-sensitive planning interventions. For example, the World Health Organization's Age-Friendly Cities Framework ([World Health Organization, 2007](#)) encourages cities to adapt the built environment and services to the diverse needs and capacities of older residents.

From a planning perspective, planning departments can use the built-environment characteristics associated with different socio-demographic groups to further identify differences in their built environments, develop place-specific environmental improvement policies, and provide more targeted evidence for subsequent site assessment, public space improvement, and community renewal. For instance, retired professionals are typically associated with vegetation. Planning departments could pay greater attention to environmental elements such as greenery and consider these characteristics during community renewal. For Ethnically Diverse Suburban Professionals (Group 5), residential scene features make a relatively high contribution to classification, indicating that the residential environment is an important visual dimension for distinguishing this group. Planning and renewal in these communities could further consider built-environment characteristics such as housing form, street space, and the surrounding public environment. Therefore, SVI-based geodemographic classification not only facilitates understanding of the living environments of different socio-demographic groups but also informs improvements to public spaces and supports the environmental needs of diverse populations.

5.5. Limitations and future work

This study has some limitations. First, it lacks an analysis of the temporal evolution of the visual environment, making this static perspective difficult to capture the dynamics of urban development. Future work should collect streetscape data from different time periods and analyse the temporal changes in visual features ([Li, 2021](#); [Liang et al., 2023](#)) to reveal the impact of these changes on the transfer of socio-demographic groups. Second, the model relies exclusively on visual features and does not incorporate socioeconomic dimensions such as financial precarity ([Ye and Singleton, 2026](#)), which may influence or mediate the distribution patterns of socio-demographic groups. While the sole reliance on SVI is one of the strengths of the work, incorporating these additional data sources in future research could yield a more comprehensive understanding of the factors influencing group distributions and enhance the robustness of socio-demographic analyses ([Singleton and Spielman, 2026](#)). Third, SVI coverage, update frequency and image capture conditions, such as time, weather and lighting, may be uneven across cities and neighbourhoods, which may lead to spatial, temporal and visual bias ([Kim and Jang, 2023](#); [Fan et al., 2025](#); [Zhao et al., 2025](#); [Hou et al., 2024](#)). Fourth, although the proposed framework is scalable,

the current analysis is conducted within the England and Wales context. Therefore, the findings may not fully capture how transfer bias manifests in other countries, where socio-demographic structures and visual environments may differ substantially (Goodchild and Li, 2021). Future research could extend the framework to cities in other countries to examine whether transfer bias follows similar patterns across national contexts and whether the transferability of SVI-based geodemographic inference observed in England and Wales can also be found in other countries. Fifth, as an area-level geodemographic classification, the OAC may obscure population heterogeneity within individual OAs, as well as differences between areas assigned to the same category. Therefore, inconsistencies between SVI-based predictions and OAC labels may arise from limited visual distinguishability, within-category heterogeneity, local variation in visual-demographic relationships, and the applicability of national OAC categories across different local contexts. Future research could compare the national OAC with locally specific classifications, such as the London-specific OAC (Singleton and Longley, 2024), and incorporate field validation to further distinguish between local variation in visual-demographic relationships and limitations arising from the application of a national classification across diverse local contexts.

6. Conclusion

This study proposes a new approach for inferring and understanding socio-demographic groups using street view imagery, a widely available spatial data source across the world but underutilized for this purpose. By treating multi-dimensional visual features as indicators of the socio-demographic distribution of the population, the study systematically demonstrates the promising potential of SVI to capture the characteristics of socio-demographic groups. In addition, we develop a scalable model showing that, even in the absence of complete socio-statistical data, partial samples can still yield reliable predictions for unobserved areas. The proposed framework provides new perspectives for incorporating SVI into research on geodemographic classification. This approach offers planners and local authorities a detailed and scalable framework to develop place-specific environmental design policies that address the needs of diverse populations and support inclusivity and well-being in urban contexts.

Our results reveal that multi-dimensional visual features can effectively predict the spatial distribution of socio-demographic groups. Different groups demonstrate significant disparities in the built environment features associated with their residential contexts. Retired Professionals are more associated with visual features such as vegetation, whereas residential scene features are more typical of Ethnically Diverse Suburban Professionals. These results indicate that integrating SVI into geodemographic classification provides additional street-level insights that complement traditional census data. From an urban planning perspective, this environmental information facilitates the identification of differences in neighbourhood built-environment characteristics among various socio-demographic groups, thereby providing further evidence for site assessment, public-space

enhancement, and neighbourhood renewal. Furthermore, cross-city analyses demonstrate that visual–demographic relationships differ across urban contexts, highlighting the necessity of considering local environmental characteristics when applying national geodemographic patterns to specific locations.

Acknowledgments

We thank our colleagues at the NUS Urban Analytics Lab for their valuable discussions and insights. This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295-000-171-133. This research is part of the project Time series Street View Imagery - A new paradigm for urban sensing, which is supported by the Singapore Ministry of Education Academic Research Fund Tier 1. The second author is supported by the China Scholarship Council.

CRediT authorship contribution statement

K.Z.: Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization. **Y.C.:** Conceptualization, Methodology, Writing - Review & Editing. **P.L.:** Conceptualization, Writing - Review & Editing, Supervision. **Y.H.:** Writing - Review & Editing. **F.B.:** Conceptualization, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT 5.4 to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Appendix A. Features Summary

Table A.4 shows all the image-derived variables used in this study.

Table A.4: Summary of all the image-derived variables

Visual feature categories	Image-derived variables summary
Semantic	• 38 semantic categories: – Bicycle, building, bus, car, fence, motorcycle, person, pole, rider, road, sidewalk, sky, terrain, traffic light, traffic sign, train, truck, vegetation, wall. Mean and standard deviation of 19 features.
Object	• 32 object categories:

Visual feature categories	Image-derived variables summary
Scene	– Person, building, road, bicycle, car, tree, river, wall, streetlight, trash can, barrier, crosswalk, pole, sidewalk, flower, grass, parking meter, motorbike, fence, aeroplane, bus, train, truck, boat, traffic light, traffic sign, fire hydrant, stop sign, sky, bench, bird, dog. • 473 scene-related features: – 365 scene categories, a list of categories can be found at: https://github.com/CSAILVision/places365/blob/master/categories_places365.txt – 102 scene attributes, a list of attributes can be found at: https://github.com/CSAILVision/places365/blob/master/labels_sunattribute.txt – 6 scene diversity indices: the Shannon and Simpson diversity indices of the scene categories and attributes in the parcel, these two indices weighted by the parcel area, and these two indices per square kilometre of the parcel
Pixel	• 9 low-level visual features: – Canny, laplacian, sobelX, sobelY, hue, lightness, saturation, blob_count, mean_depth
Spatial	• 4 neighbourhood diversity features: – The mean area-weighted Shannon and Simpson diversity of adjacent parcels – The mean Shannon and Simpson diversity per square kilometer of adjacent parcels

Appendix B. SVI sampling density across cities

Figure B.12 further illustrates the distribution of SVI sampling density across different OAs in each city. The sampling density is calculated as:

$$D_i = \frac{N_i}{A_i}, \quad (\text{B.1})$$

where D_i denotes the SVI sampling density of OA i (points/km²), N_i denotes the number of SVI samples within OA i , and A_i denotes the area of OA i in km².

The H3 spatial indexing system was used to generate SVI sampling points, resulting in an initially even spatial distribution of sampling locations across the study areas. However, since GSV imagery is primarily collected along roads accessible to vehicles, certain residential roads and other regions may lack valid imagery. Therefore, the actual distribution of SVI may diverge from the initial H3-based distribution.

To assess this variation, we calculated SVI sampling density per unit area for each OA. Variations in OA area, shape, and the number of available images led to differences in sampling densities. Notably, in smaller OAs, even a single or a few SVI samples can produce a high sampling density per unit area. To avoid a small number of extreme values

compressing the visible range of the main distribution, the figure displays only observations with sampling densities within the mean plus or minus two standard deviations ($[\mu - 2\sigma, \mu + 2\sigma]$) for each city.

The results indicate that the median sampling densities of the ten cities are generally consistent, ranging from approximately 62 to 69 points per square kilometre. This suggests that, under a consistent H3 resolution and sampling strategy, typical SVI sampling intensity across OAs is broadly comparable among different cities. Remaining variations within and between cities may be influenced by factors such as road network density, GSV availability, and OA area and shape. Future research should consider integrating multiple street-view data sources, including Mapillary and KartaView, to enhance spatial coverage and mitigate the effects of coverage disparities associated with a single data source.

Appendix C. Model performance

Appendix C.1. Weighted-F1 of local models

Table C.5 shows the Weighted-F1 scores of various city-level models on different machine learning models. Combining the prediction accuracy of each city model with the accuracy and weighted-F1 of the national model further confirms that the classification model based on visual features can effectively infer socio-demographic groups.

Table C.5: Comparison of model performance in 10 UK cities using different machine learning algorithms under two training stages. Evaluation metrics: Weighted-F1. Bold numbers indicate the highest Weighted-F1 achieved for each city across all models.

Model	Number of training sessions	Birmingham	Bristol	Cardiff	Leeds	Liverpool	London	Manchester	Nottingham	Sheffield	Southampton
CatBoost	L1	0.5643	0.4756	0.4580	0.4625	0.4591	0.6577	0.5667	0.5654	0.4131	0.3986
	L2	0.6052	0.499	0.4297	0.4832	0.4616	0.6678	0.5903	0.586	0.457	0.4122
Extra Trees	L1	0.5044	0.4341	0.3781	0.3642	0.4	0.6125	0.5346	0.5689	0.363	0.4045
	L2	0.5898	0.4843	0.4283	0.4721	0.4333	0.6639	0.5854	0.5777	0.4353	0.413
LightGBM	L1	0.5865	0.4764	0.478	0.4474	0.4616	0.6662	0.5853	0.5841	0.4194	0.4426
	L2	0.5999	0.491	0.4779	0.4692	0.47	0.6735	0.6091	0.5889	0.4707	0.4413
NeuralNetFastAI	L1	0.5931	0.5162	0.4554	0.4729	0.4558	0.6696	0.6041	0.5899	0.4428	0.4762
	L2	0.5881	0.5251	0.4203	0.4564	0.4339	0.6668	0.6019	0.5925	0.4481	0.4974
NeuralNetTorch	L1	0.5893	0.5367	0.4892	0.4785	0.4794	0.6646	0.6122	0.6266	0.4714	0.5276
	L2	0.6034	0.521	0.4752	0.4635	0.4581	0.6745	0.6154	0.6254	0.4714	0.5043
Random Forest	L1	0.5407	0.4382	0.4073	0.423	0.4361	0.6325	0.5306	0.5469	0.3947	0.4138
	L2	0.5987	0.4946	0.4575	0.4826	0.475	0.6663	0.5959	0.576	0.4727	0.4285
XGBoost	L1	0.5816	0.4721	0.4677	0.4298	0.4392	0.6636	0.5554	0.5875	0.4078	0.4195
	L2	0.6001	0.4985	0.4495	0.4716	0.461	0.6714	0.5876	0.5816	0.4763	0.4713

Appendix C.2. National model performance

Table C.6 shows the performance of the national model trained with the LightGBM algorithm. We also provide the number of samples for each socio-demographic group and their corresponding proportions. Notably, some socio-demographic groups have relatively

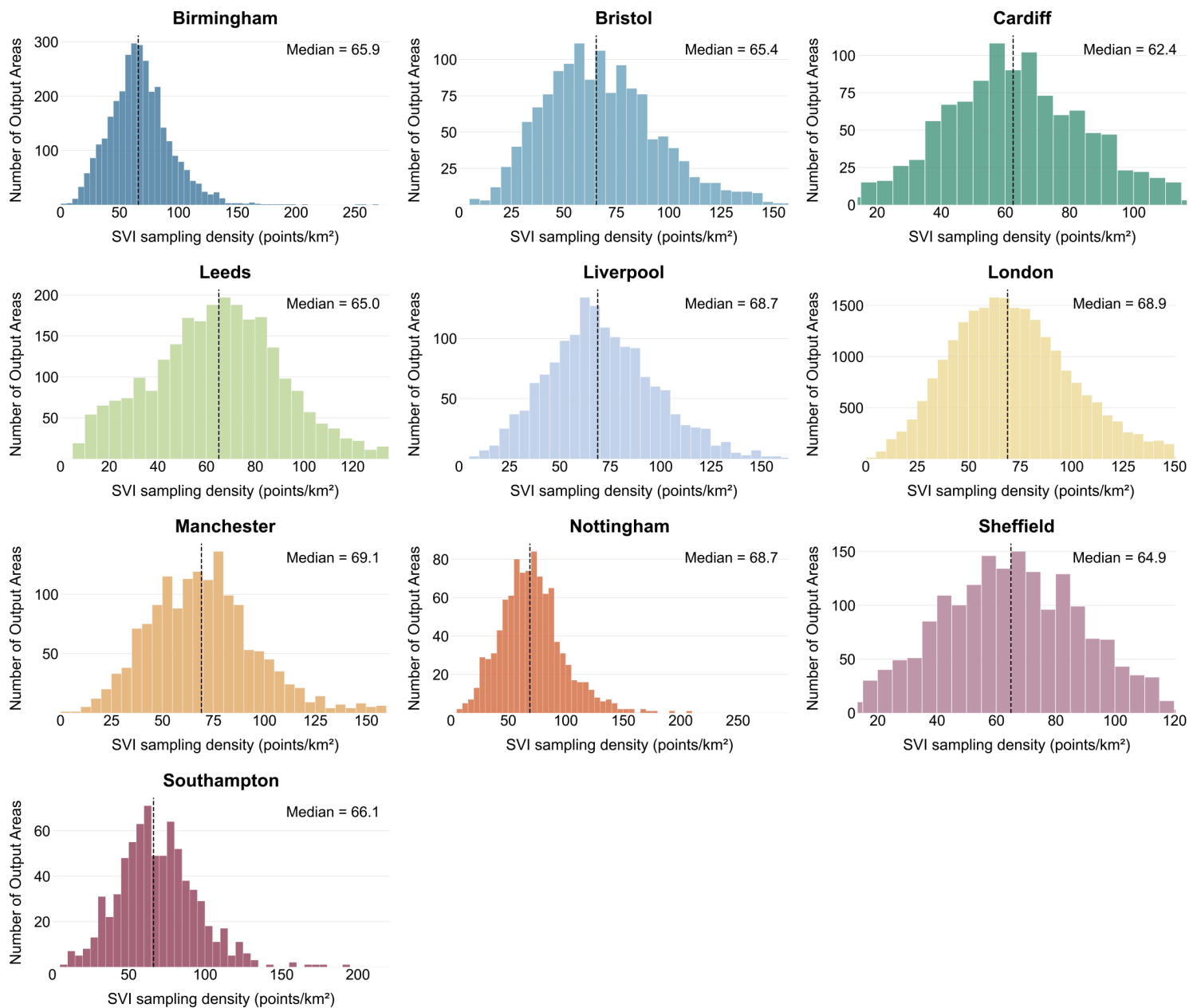


Figure B.12: Distribution of OA-level SVI sampling density across ten cities.

Table D.7: Sample size for each socio-demographic group in the studied England and Wales cities.

City	Total	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
Birmingham	3233	57	65	196	1542	676	629	40	28
Bristol	1368	35	108	213	165	119	639	66	23
Cardiff	1090	111	115	120	139	181	369	44	11
Leeds	2560	303	479	184	373	396	532	219	74
Liverpool	1547	133	149	149	127	99	470	358	62
London	23386	125	119	8241	8581	3601	2695	9	15
Manchester	1527	10	15	248	725	164	347	7	11
Nottingham	983	18	17	103	344	169	295	25	12
Sheffield	1804	267	275	103	265	169	444	217	64
Southampton	754	17	71	106	95	118	269	38	40

small sample sizes, the model learns the limited information during training and leads to poorer performance for these socio-demographic groups.

Table C.6: National model performance across all socio-demographic groups and sample sizes of each group.

Category	Total	F1-score	Precision	Recall
Retired Professionals (1)	1076 (2.81%)	0.3270	0.5049	0.2419
Suburbanites and Peri-Urbanites (2)	1413 (3.69%)	0.3041	0.4371	0.2332
Multicultural and Educated Urbanites (3)	9663 (25.26%)	0.7960	0.7729	0.8205
Low-Skilled Migrant and Student Communities (4)	12356 (32.30%)	0.6596	0.6118	0.7155
Ethnically Diverse Suburban Professionals (5)	5692 (14.88%)	0.6185	0.5920	0.6476
Baseline UK (6)	6689 (17.49%)	0.3987	0.4333	0.3692
Semi- and Un-Skilled Workforce (7)	1023 (2.67%)	0.2691	0.5286	0.1805
Legacy Communities (8)	340 (0.89%)	0	0	0

Appendix D. Sample size of all the cities

As shown in the table D.7. We present the sample sizes for all cities and the sample size of OAC groups within each city.

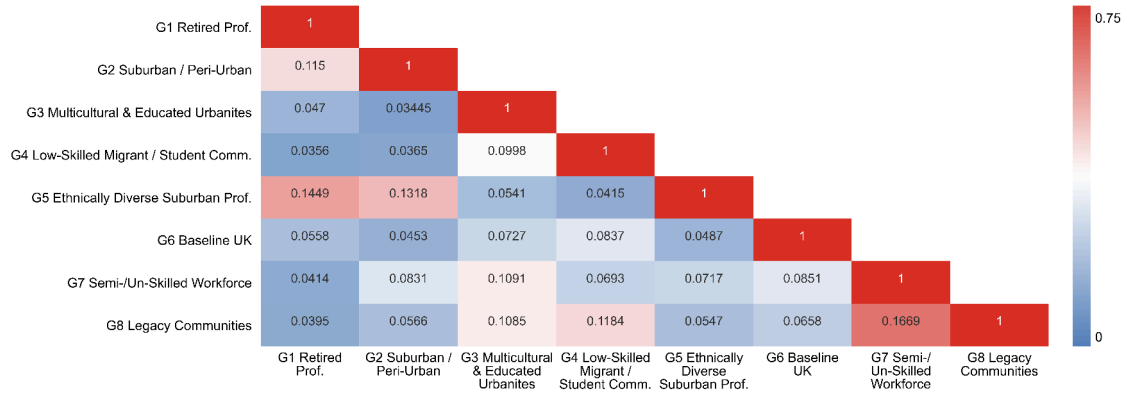


Figure E.13: Cosine similarity of different socio-demographic groups definitions

Appendix E. Cosine similarity

As shown in Figure E.13, we use cosine similarity to assess the similarity between the official definitions of the socio-demographic groups. Variations in classification performance partially reflect the degree of separability among these groups. When groups possess similar social attributes, the maximum achievable discrimination is limited by the overlap in their underlying data distributions. For instance, Groups 1 (Retired Professionals) and 5 (Ethnically Diverse Suburban Professionals) show one of the highest cosine similarities among all group pairs (0.1449). Group 1 also shows low classification performance in the national model, and its samples are frequently predicted as Group 5 in cross-city transfer (Figure G.16). These findings suggest that some of the model’s misclassifications may be attributed to overlapping social attributes across groups, such as occupation, age, and educational level, which contribute to shared environmental contexts in daily life.

Appendix F. Transfer bias results

Figure F.14 presents the cross-city transfer results for the 10 cities selected for this study. The results indicate that the transfer bias generated when applying a trained model to other cities can serve as an effective metric for quantifying urban spatial heterogeneity. In addition, we observe certain consistent spatial patterns across cities. For example, Group 3 (Multicultural and Educated Urbanites) shows strong performance in both prediction and transfer across all 10 cities. On the one hand, Group 3 accounts for a relatively large proportion of the sample in several cities (please refer to Appendix D for sample sizes), enabling the model to learn more stable feature representations. On the other hand, this group is typically concentrated in the high-density core areas of cities (e.g., the City of London and central districts of other major metropolitan areas), where built environment

characteristics, such as dense building fabric and commercial facades, tend to converge across cities, resulting in more consistent visual patterns.

In contrast, Group 1 (Retired Professionals) and Group 8 (Legacy Communities) exhibit weaker transfer performance in most cities (Figure F.14 and Figure G.16). For example, when transferring models from Liverpool, London, and Manchester to Leeds, the prediction accuracy for Group 1 declines substantially. Similarly, when transferring models from Birmingham, Leeds, and Nottingham to Southampton, the prediction performance for Group 8 drops noticeably. Notably, transferability is generally higher when models are applied to London and Manchester, suggesting that the visual features of the training and testing cities exhibit greater alignment in these cases.

Collectively, these cross-city transfer results not only evaluate the generalizability of the model but also capture the extent to which cities differ in their visual environments, providing empirical evidence of urban visual heterogeneity.

To better understand the wrong prediction, we further spatially visualize the transfer results using binary maps (Figure F.15). Compared to applying models from other cities to the London dataset, the number of misclassifications is the highest when transferring Liverpool's model to London, and combined with Figure 9 transfer accuracy, the transfer accuracy is also the lowest. Further inference can be made by combining the socioeconomic data that the number of immigrants in London is much higher than that in Liverpool (Table H.8), indicating that its demographic composition may be more complex, and when transferring the model from Liverpool to London, the model learns limited information, and is more likely to map the more complex neighbourhoods in the testing city to the more homogeneous composition of the neighbourhoods in the training city, resulting in more misclassified areas. In addition, Birmingham and Bristol are more similar in terms of population density, but differ more in the number of migrants and the number of people not working. When these two cities transfer to each other, there are more misclassification areas, suggesting that the labour force structure and population mobility of the cities may also affect group label boundaries under conditions of similar population densities, which may reduce the transferability of the model.

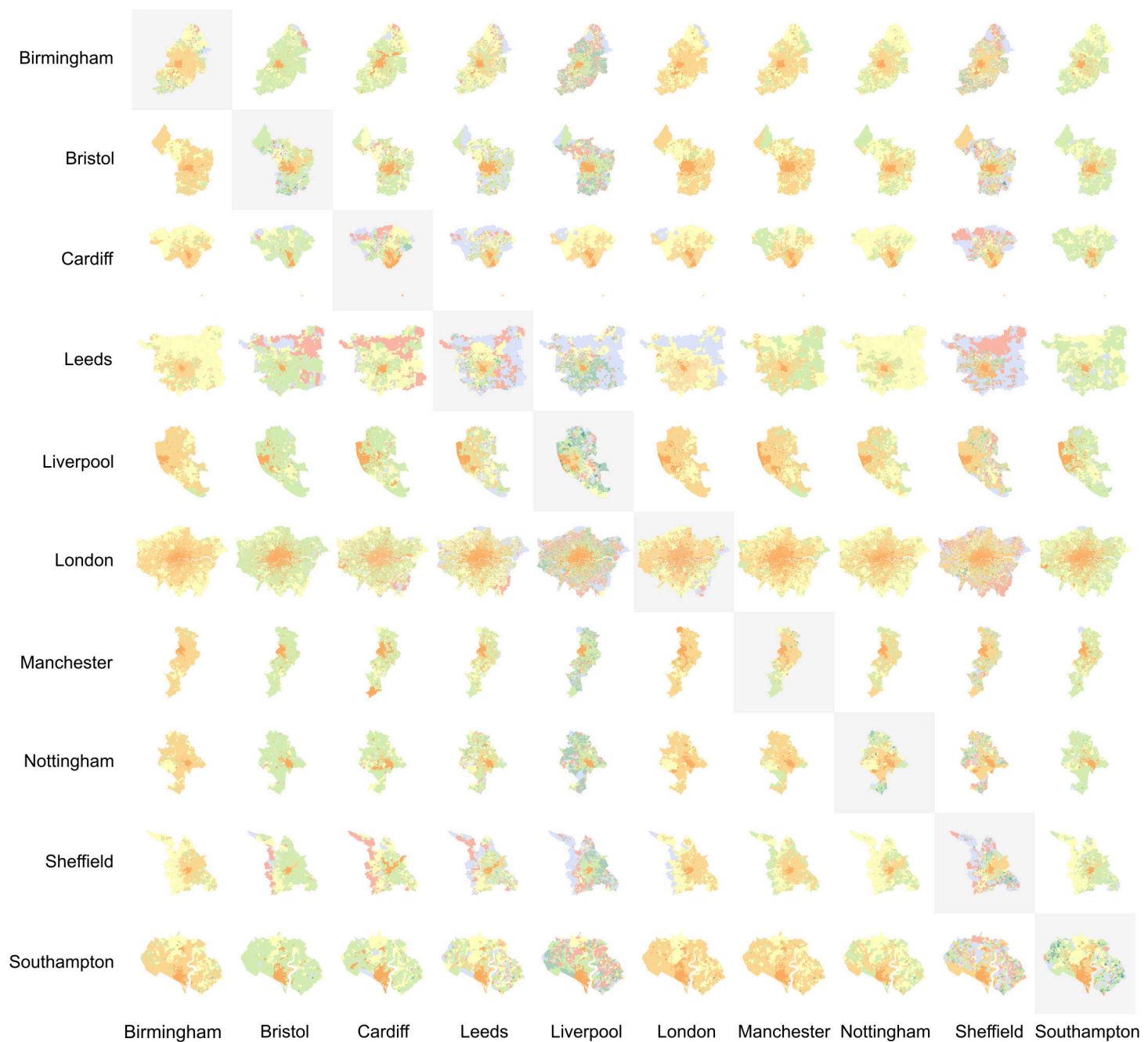


Figure F.14: Transfer analysis results. The horizontal axis represents training cities, and the vertical axis represents test cities.

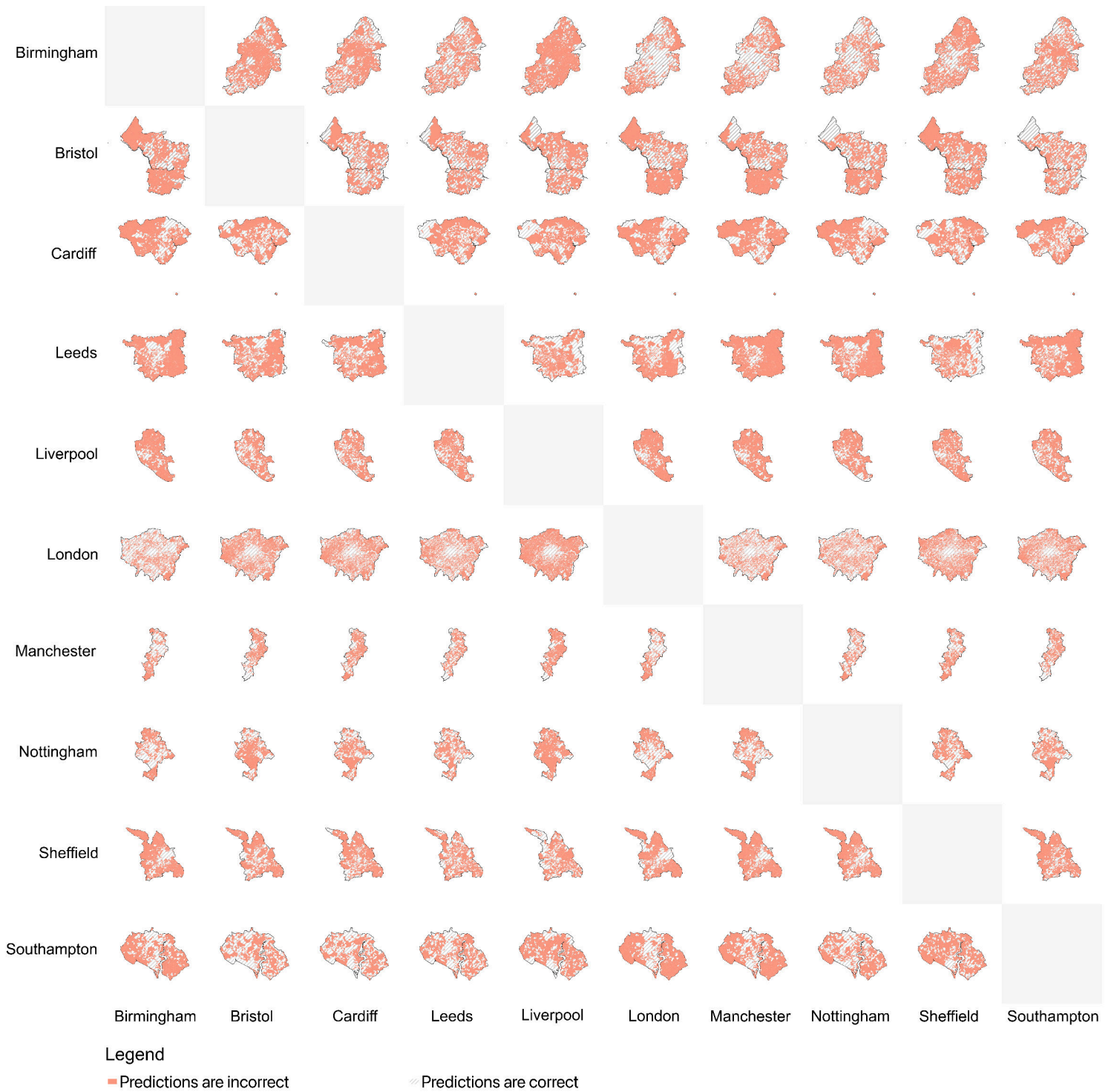


Figure F.15: Binary map of city-level model transfer results. The horizontal axis represents training cities, and the vertical axis represents test cities.

Appendix G. Confusion matrices for cross-city model transfer

Figure G.16 illustrates the prediction patterns of various OAC categories during cross-city model transfer. Each panel represents a training city, with the corresponding model evaluated on the remaining nine cities; their predictions were then aggregated into a confusion matrix and normalized by the ground truth class. The diagonal elements indicate the proportion of each OAC category correctly identified after cross-city transfer, while the off-diagonal elements indicate misclassification patterns between categories. These findings reveal differences in the transfer performance of OAC categories across cities, as well as potential visual overlap and variation in distinguishability between categories.



Figure G.16: Class-level confusion matrices for cross-city model transfer

Appendix H. Socio-economic data

Table H.8 presents selected socioeconomic indicators for each city.

Table H.8: City-level indicators (Census 2021).

City	Density (persons/km ²)	Migrant indicator (outside the UK, Units:persons)	Never worked (Unemployment history, Units:persons)	Economically active and a full-time student: In employment (Economic activity status, Units:persons)
Birmingham	4,275.4	13,642	186,119	19,087
Bristol	4,308.1	6,664	39,291	9,726
Cardiff	2,571.3	5,181	38,809	8,157
Leeds	1,471.7	9,592	77,744	14,856
Liverpool	4,346.1	7,241	66,625	10,841
London	5,597.6	149,532	966,729	117,127
Manchester	4,772.7	12,659	81,350	12,896
Nottingham	4,337.6	6,245	43,571	8,700
Sheffield	1,512.5	6,709	63,214	11,243
Southampton	4,990.5	4,965	24,069	5,520

References

- Anderson, T.K., 2010. Using geodemographics to measure and explain social and environment differences in road traffic accident risk. *Environment and Planning A* 42, 2186–2200.
- Barber, A., Hall, S., 2008. Birmingham: Whose urban renaissance? Regeneration as a response to economic restructuring. *Policy Studies* 29, 281–292.
- Biljecki, F., Ito, K., 2021. Street view imagery in urban analytics and gis: A review. *Landscape and Urban Planning* 215, 104217.
- Bruch, E.E., 2014. How population structure shapes neighborhood segregation. *American Journal of Sociology* 119, 1221–1278.
- Chappell, P., Tse, M., Zhang, M., Moore, S., 2017. Using gps geo-tagged social media data and geodemographics to investigate social differences: A twitter pilot study. *Sociological Research Online* 22, 38–56.
- Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R., 2022a. Masked-attention mask transformer for universal image segmentation, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1290–1299.
- Cheng, T., Liu, C., Yang, H., Wang, N., Liu, Y., 2022b. From service capacity to spatial equity: Exploring a multi-stage decision-making approach for optimizing elderly-care facility distribution in the city centre of tianjin, china. *Sustainable Cities and Society* 85, 104076.
- Coast, E., Fanghanel, A., Lelievre, E., Randall, S., 2016. Counting the population or describing society? a comparison of english and welsh and french censuses. *European Journal of Population* 32, 165–188.
- Crewe, L., Beaverstock, J., 1998. Fashioning the city: Cultures of consumption in contemporary urban spaces. *Geoforum* 29, 287–308.
- Dabinett, G., 2005. Competing in the information age: Urban regeneration and economic development practices in the city of sheffield, united kingdom. *Journal of Urban Technology* 12, 19–38.
- De Sabbata, S., Liu, P., 2023. A graph neural network framework for spatial geodemographic classification. *International Journal of Geographical Information Science* 37, 2464–2486.
- Du, T., Fertner, C., Jiang, W., Andersen, L.M., Vejre, H., 2023. Understanding the change in the social networks of residential groups affected by urban renewal. *Environmental Impact Assessment Review* 98, 106970.

Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., Smola, A., 2020. Autogluon-tabular: Robust and accurate automl for structured data. arXiv preprint arXiv:2003.06505 .

Ethnicity facts and figures, 2022. Regional ethnic diversity. <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/regional-ethnic-diversity/latest/>.

Fan, Z., Feng, C.C., Biljecki, F., 2025. Coverage and bias of street view imagery in mapping the urban environment. Computers, Environment and Urban Systems 117, 102253. doi:10.1016/j.compenvurbsys.2025.102253.

Fan, Z., Zhang, F., Loo, B.P., Ratti, C., 2023. Urban visual intelligence: Uncovering hidden city profiles with street view images. Proceedings of the National Academy of Sciences 120, e2220417120.

Gale, C.G., Singleton, A., Bates, A.G., Longley, P.A., 2016. Creating the 2011 area classification for output areas (2011 oac). Journal of Spatial Information Science 12, 1–27.

Gao, Q.L., Yue, Y., Tu, W., Cao, J., Li, Q.Q., 2021. Segregation or integration? exploring activity disparities between migrants and settled urban residents using human mobility data. Transactions in GIS 25, 2791–2820.

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E.L., Fei-Fei, L., 2017. Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. Proceedings of the National Academy of Sciences 114, 13108–13113.

Goodchild, M.F., Li, W., 2021. Replication across space and time must be weak in the social and environmental sciences. Proceedings of the National Academy of Sciences 118, e2015759118.

Gulma, U.L., 2022. A new geodemographic classification of the influence of neighbourhood characteristics on crime: The case of leeds, uk. Computers, Environment and Urban Systems 92, 101748.

Hedman, L., 2011. The impact of residential mobility on measurements of neighbourhood effects. Housing studies 26, 501–519.

Hipp, J.R., 2020. Neighborhood change from the bottom up: What are the determinants of social distance between new and prior residents? Social science research 86, 102372.

- Hou, C., Zhang, F., Kang, Y., Gao, S., Li, Y., Duarte, F., Li, S., 2025. Transferred bias uncovers the balance between the development of physical and socioeconomic environments of cities. *Annals of the American Association of Geographers* 115, 148–166.
- Hou, Y., Quintana, M., Khomiakov, M., Yap, W., Ouyang, J., Ito, K., Wang, Z., Zhao, T., Biljecki, F., 2024. Global streetscapes—a comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics. *ISPRS Journal of Photogrammetry and Remote Sensing* 215, 216–238.
- Hu, S., Zhang, Z., Ge, D., Lin, X., Lu, Y., 2024. Residential differentiation characteristics based on “socio-spatial” coupling: A case study of zhengzhou. *Applied Geography* 166, 103269.
- Ito, K., Zhu, Y., Abdelrahman, M., Liang, X., Fan, Z., Hou, Y., Zhao, T., Ma, R., Fujiwara, K., Ouyang, J., et al., 2025. Zensvi: An open-source software for the integrated acquisition, processing and analysis of street view imagery towards scalable urban science. *Computers, Environment and Urban Systems* 119, 102283.
- Kang, Y., Zhang, F., Gao, S., Lin, H., Liu, Y., 2020. A review of urban physical environment sensing using street view imagery in public health studies. *Annals of GIS* 26, 261–275.
- Kim, J., Jang, K.M., 2023. An examination of the spatial coverage and temporal variability of google street view (gsv) images in small-and medium-sized cities: A people-based approach. *Computers, Environment and Urban Systems* 102, 101956.
- Leeds City Council, 2022. Authority Monitoring Report 2019–21. Technical Report. Leeds City Council. URL: <https://www.leeds.gov.uk/sites/default/files/docs/Authority%20Monitoring%20Report%202019-21.pdf>.
- Li, X., 2021. Examining the spatial distribution and temporal change of the green view index in new york city using google street view images and deep learning. *Environment and Planning B: Urban Analytics and City Science* 48, 2039–2054.
- Li, X., Zhang, C., Li, W., Kuzovkina, Y.A., Weiner, D., 2015. Who lives in greener neighborhoods? the distribution of street greenery and its association with residents’ socioeconomic conditions in hartford, connecticut, usa. *Urban Forestry & Urban Greening* 14, 751–759.
- Liang, X., Zhao, T., Biljecki, F., 2023. Revealing spatio-temporal evolution of urban visual environments with street view imagery. *Landscape and Urban Planning* 237, 104802.
- Lin, Y., 2024. You are where you live? evaluating the racial and ethnic (mis) representation in geodemographic classification. *Applied Geography* 165, 103244.

- Lipovac, N., 1997. Space and place. *Prostor* 5, 1–34.
- Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., et al., 2024. Grounding dino: Marrying dino with grounded pre-training for open-set object detection, in: *European conference on computer vision*, Springer. pp. 38–55.
- Luo, E., Kuffer, M., Wang, J., 2022. Urban poverty maps-from characterising deprivation using geo-spatial data to capturing deprivation from space. *Sustainable Cities and Society* 84, 104033.
- Ma, R., Wang, W., Zhang, F., Shim, K., Ratti, C., 2019. Typeface reveals spatial economical patterns. *Scientific Reports* 9, 15946.
- Mavrogeni, M., Cheshire, J., Iliev, M., van Dijk, J., 2025. Creating the London night-worker geodemographic classification. *Environment and Planning B: Urban Analytics and City Science* 53, 1769–1793.
- Mavrogeni, M., Longley, P.A., van Dijk, J.T., 2026. Geodemographic classification of interactivity: Proof of concept for greater london. *Annals of the American Association of Geographers* , 1–21.
- Meegan, R., 2015. City profile–leeds. *Cities* 42, 42–53.
- Ministry of Housing, Communities and Local Government, 2024. National Planning Policy Framework. Technical Report. HM Government. London. URL: <https://www.gov.uk/guidance/national-planning-policy-framework>.
- Nagendra, H., 2002. Opposite trends in response for the shannon and simpson indices of landscape diversity. *Applied geography* 22, 175–186.
- Office for National Statistics, 2022. Census 2021 demography and international migration statistics. <https://www.ons.gov.uk/news/news/census2021demographyandinternationalmigrationstatistics>.
- Office for National Statistics, 2026. Quarterly update on population and migration statistics. URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/quarterlyupdateonpopulationandmigrationstatistics/latest>.
- O'Regan, A.C., Hunter, R.F., Nyhan, M.M., 2021. “biophilic cities”: Quantifying the impact of google street view-derived greenspace exposures on socioeconomic factors and self-reported health. *Environmental science & technology* 55, 9063–9073.

- Peck, J., Ward, K. (Eds.), 2002. *City of revolution: Restructuring Manchester*. Manchester University Press, Manchester.
- Petersen, J., Gibin, M., Longley, P., Mateos, P., Atkinson, P., Ashby, D., 2011. Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *Journal of Geographical Systems* 13, 173–192.
- Pike, A., Béal, V., Cauchi-Duval, N., Franklin, R., Kinossian, N., Lang, T., Leibert, T., MacKinnon, D., Rousseau, M., Royer, J., et al., 2024. ‘left behind places’: a geographical etymology. *Regional Studies* 58, 1167–1179.
- Pinch, S., 2002. Southampton. *Cities* 19, 71–78.
- Ranftl, R., Bochkovskiy, A., Koltun, V., 2021. Vision transformers for dense prediction, in: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12179–12188.
- Rossetti, T., Lobel, H., Rocco, V., Hurtubia, R., 2019. Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach. *Landscape and urban planning* 181, 169–178.
- Sassen, S., 2013. *The global city: New york, london, tokyo* .
- Batista e Silva, F., Freire, S., Schiavina, M., Rosina, K., Marín-Herrera, M.A., Ziemba, L., Craglia, M., Koomen, E., Lavallo, C., 2020. Uncovering temporal changes in europe’s population density patterns using a data fusion approach. *Nature communications* 11, 4631.
- Singleton, A., Alexiou, A., Savani, R., 2020. Mapping the geodemographics of digital inequality in Great Britain: An integration of machine learning into small area estimation. *Computers, Environment and Urban Systems* 82, 101486.
- Singleton, A., Spielman, S.E., 2026. Geodemographics and residential differentiation: A methodological review and future directions for learned representations of the social landscape. *Computers, Environment and Urban Systems* 125, 102396.
- Singleton, A.D., Longley, P.A., 2024. Classifying and mapping residential structure through the london output area classification. *Environment and Planning B: Urban Analytics and City Science* 51, 1153–1164.
- Song, L., Kong, X., Cheng, P., 2024. Supply-demand matching assessment of the public service facilities in 15-minute community life circle based on residents’ behaviors. *Cities* 144, 104637.

Suel, E., Polak, J.W., Bennett, J.E., Ezzati, M., 2019. Measuring social, environmental and health inequalities using deep learning and street imagery. *Scientific reports* 9, 6229.

Sykes, O., Brown, J., Cocks, M., Shaw, D., Couch, C., 2013. A city profile of liverpool. *Cities* 35, 299–318.

Tallon, A.R., 2007. Bristol. *Cities* 24, 74–88.

Tang, J., Long, Y., 2019. Measuring visual quality of street space and its temporal variation: Methodology and its application in the hutong area in beijing. *Landscape and Urban Planning* 191, 103436.

Tuan, Y.F., 1977. *Space and place: The perspective of experience*. University of Minnesota Press, Minneapolis.

Uğurel, E., Chen, C., Lee, B.H., Rodrigues, F., 2026. On predicting sociodemographics from mobility signals. *Computers, Environment and Urban Systems* 127, 102420.

UK Statistics Authority, 2025. Recommendation from the uk statistics authority on the future of population and migration statistics in england and wales. URL: <https://uksa.statisticsauthority.gov.uk/publication/recommendation-from-the-uk-statistics-authority-on-the-future-of-population-statistics/pages/3/>.

Verma, D., Jana, A., Ramamritham, K., 2020. Predicting human perception of the urban environment in a spatiotemporal urban setting using locally acquired street view images and audio clips. *Building and Environment* 186, 107340.

Vickers, D., Rees, P., 2007. Creating the uk national statistics 2001 output area classification. *Journal of the Royal Statistical Society Series A: Statistics in Society* 170, 379–403.

Waite, D., 2015. City profile: Cardiff and the shift to city-regionalism. *Cities* 48, 21–30.

Wang, R., Lu, Y., Zhang, J., Liu, P., Yao, Y., Liu, Y., 2019. The relationship between visual enclosure for neighbourhood street walkability and elders' mental health in china: Using street view images. *Journal of Transport & Health* 13, 90–102.

World Health Organization, 2007. *Global age-friendly cities: A guide*. World Health Organization.

Wu, C., Liang, Y., Zhao, M., Teng, M., Yue, H., Ye, Y., 2025. Perceiving the fine-scale urban poverty using street view images through a vision-language model. *Sustainable Cities and Society* 123, 106267.

- Wu, Q., Cheng, J., Chen, G., Hammel, D.J., Wu, X., 2014. Socio-spatial differentiation and residential segregation in the chinese city based on the 2000 community-level census data: A case study of the inner city of nanjing. *Cities* 39, 109–119.
- Wyszomierski, J., Longley, P.A., Singleton, A.D., Gale, C., O'Brien, O., 2024. A neighbourhood output area classification from the 2021 and 2022 uk censuses. *The Geographical Journal* 190, e12550.
- Yang, G., Su, Y., Zhao, M., Li, C., Wu, C., 2026. Beyond single snapshots: Quantifying multi-scale heterogeneity from street-view imagery and what it reveals about perception. *Cities* 175, 107227.
- Yang, Y., Dolega, L., Darlington-Pollock, F., 2023. Ageing in place classification: Creating a geodemographic classification for the ageing population in england. *Applied spatial analysis and policy* 16, 583–623.
- Ye, Z., Singleton, A., 2026. Thriving or surviving: Understanding the geography of financial precarity in great britain. *Computers, Environment and Urban Systems* 125, 102399.
- Yin, L., Wang, Z., 2016. Measuring visual enclosure for street walkability: Using machine learning algorithms and google street view imagery. *Applied geography* 76, 147–153.
- Zhang, F., Zhang, D., Liu, Y., Lin, H., 2018a. Representing place locales using scene elements. *Computers, Environment and Urban Systems* 71, 153–164.
- Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H.H., Lin, H., Ratti, C., 2018b. Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning* 180, 148–160.
- Zhang, G., Yang, J., Wu, G., Hu, X., 2021. Exploring the interactive influence on landscape preference from multiple visual attributes: Openness, richness, order, and depth. *Urban Forestry & Urban Greening* 65, 127363.
- Zhang, Y., Zhang, F., Fang, L., Chen, N., 2023. Inferring socioeconomic environment from built environment characteristics based street view images: An approach of seq2seq method. *International Journal of Applied Earth Observation and Geoinformation* 123, 103458.
- Zhao, T., Liang, X., Biljecki, F., Tu, W., Cao, J., Li, X., Yi, S., 2025. Quantifying seasonal bias in street view imagery for urban form assessment: A global analysis of 40 cities. *Computers, Environment and Urban Systems* 120, 102302.
- Zhao, T., Liang, X., Tu, W., Huang, Z., Biljecki, F., 2023. Sensing urban soundscapes from street view imagery. *Computers, Environment and Urban Systems* 99, 101915.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A., 2017. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence* 40, 1452–1464.