



Places for play: Understanding human perception of playability in cities using street view images and deep learning

Jacob Kruse^a, Yuhao Kang^a, Yu-Ning Liu^a, Fan Zhang^b, Song Gao^{a,*}

^a Geospatial Data Science Lab, Department of Geography, University of Wisconsin-Madison, Madison, WI 53703, United States

^b Senseable City Lab, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

ARTICLE INFO

Keywords:

Playability
Street view images
Deep learning
Built environment
Human perception of place

ABSTRACT

Play benefits childhood development and well-being, and is a key factor in sustainable city design. Though previous studies have examined the effects of various urban features on how much children play and where they play, such studies rely on quantitative measurements of play such as the precise location of play and the duration of play time, while people's subjective feelings regarding the playability of their environment are overlooked. In this study, we capture people's perception of place playability by employing Amazon Mechanical Turk (MTurk) to classify street view images. A deep learning model trained on the labelled data is then used to evaluate neighborhood playability for three U.S. cities: Boston, Seattle, and San Francisco. Finally, multivariate and geographically weighted regression models are used to explore how various urban features are associated with playability. We find that higher traffic speeds and crime rates are negatively associated with playability, while higher scores for perception of beauty are positively associated with playability. Interestingly, a place that is perceived as lively may not be playable. Our research provides helpful insights for urban planning focused on sustainable city growth and development, as well as for research focused on creating nourishing environments for child development.

1. Introduction

Play is a crucial component of childhood growth and well-being. Children of all ages develop their abilities to interact with others and practice their social skills as they play, while learning resilience and confidence in the process (Ginsburg et al., 2007). Parents and children bond and learn to communicate when they play together in their neighborhoods (Ginsburg et al., 2007; Henry, 1990; Tamis-LeMonda, Shannon, Cabrera, & Lamb, 2004). Regarding physical and mental health, children develop and maintain healthy minds and bodies through play. Increased physical activity has also been associated with a lower prevalence of obesity in children (McCurdy, Winterbottom, Mehta, & Roberts, 2010; Summerbell et al., 2005). Such physical activity also helps build and maintain healthy bones and muscles, and has been associated with greater coordination and strength, as well a lower body mass index (Lämmle, Ziegler, Seidel, Worth, & Bös, 2013).

Given the myriad benefits of play for children's growth and development, many researchers have studied where and how children play. In a seminal work *The Death and Life of Great American Cities* (1961), Jacobs

identifies sidewalks as the central location of play, as sidewalks at the time enhanced safety, contact, and the assimilation of children into a place (Jacobs, 2016). From this framework, play arises from a variety of attributes of a given place, not merely from the presence of a single play feature such as a playground. Hence, it is necessary to construct places that are conducive to play for children. Designing such spaces in cities also meets the 11th Sustainable Development Goal (SDG), as outlined by the United Nations, to "make cities and human settlements inclusive, safe, resilient and sustainable".¹ Such a goal provides a series of measurable targets and blueprints for future cities, including universal access to basic public services. In particular, children need playable places that are accessible and safe. Therefore, examining and designing playable places benefits sustainable city development (Duarte & Alvarez, 2021).

Societal changes throughout the past several decades, however, have led to changes in where and how children play.

In rural areas, lack of transportation, lack of nearby friends and things to do, and the guarding of privately-owned natural areas discourage play (Spencer, 2006). In urban areas, lack of maintenance in

* Corresponding author.

E-mail address: song.gao@wisc.edu (S. Gao).

¹ <https://sdgs.un.org/goals>

parks and playgrounds, boringly-safe play equipment, and ever-increasing traffic speeds reduce how much children play (Wridt, 2010).

Though more prevalent in families where the parents are highly educated, parents today are generally more concerned about safety than in previous generations (Ginsburg et al., 2007). As a result, they are more likely to schedule activities for their children, relegating play to organized activities that are frequently indoors and deprive children of important contact with nature. Play in natural areas and green spaces provides both physical and mental benefits. Recent studies have found that proximity to a variety of natural features is associated with lower body mass index (BMI), suggesting that greater physical activity occurs in such areas (Dadvand et al., 2014; Lee, Heo, Jayaraman, & Dawson, 2019). Features such as trees, gardens, and nature trails in “green” school grounds have been shown to encourage play in a wider variety of students when compared to schools with only grass and concrete, since such natural features allow for various forms of non-competitive exercise (Dyment & Bell, 2008). In two studies using remotely sensed imagery to calculate a greenness index, greater greenery (e.g. more trees and grass) was associated with lower BMI in children living nearby, suggesting greater physical activity (Bell, Wilson, & Liu, 2008; Grigsby-Toussaint, Chi, & Fiese, 2011). For mental health, play in and even just passive exposure to natural areas has been shown to improve psychological health in children (McCurdy et al., 2010). Taken all together, there are many reported benefits of such natural areas for children.

In summary, a variety of attributes of the built environment and society contribute to how and where children play. We consider such attributes as affordances of a place which contribute to its “playability”. Increasing place playability, then, allows children to receive more of the benefits of play as individuals. We synthesize this conceptual framework in Fig. 1.

Within the conceptual framework shown in Fig. 1, various aspects of the built environment contribute to playability. Therefore, studying the built environment is a critical part of understanding how to create playable places. Having a method for scoring or evaluating playability itself, therefore, is also necessary. Recently, research focused on characterizing the built environment has flourished due to the availability of rich data sources, such as street view images, and recent advances in deep learning models. In particular, such models have been used to learn the features in street view images that are associated with various aspects of human perception, and produce accurate predictions of human perception on unseen images. This methodology is more scalable than traditional surveys that link human perception to the built environment, allowing researchers to study human perception at the scale of cities and larger. Furthermore, this methodology allows human perception to be what is evaluated, rather than using proxies for human perception. In the case of playability, for example, human perception of place

playability could be quantitatively predicted at scale by a trained deep learning model, rather than using a proxy for place playability such as playground count.

The ability of such deep learning models to predict human perception of street view images has two particular advantages in the fields of Geography and Urban Planning. First, the geographic location of the street view images evaluated by the model can be used to produce maps of human perception across cities, which can provide intuitive understanding as to the spatial nature of the phenomena being studied. Second, the model predictions and their associated locations can be used in geographic regression models, which allow researchers to gain insights into how various features contribute to human perception, and how their relative contributions vary across space.

In this paper, we propose a framework that quantitatively and comprehensively measures the playability of neighborhoods and examines its determining geographic and environmental factors. Within this framework, we ask the following two research questions: 1) how can playability scores be generated for neighborhoods in different cities across the United States? Is harnessing crowd-sourced labelled street view images viable as a new data source for evaluating human perception of playability? 2) what features are associated with playability and how do their effects change across space? By using multi-source big data such as street view images along with deep learning methods, we are able to take full advantage of all of the features visible in street view images.

Regarding the use of deep learning models and street view images to audit the urban environment, our contribution in this work is twofold: First, our research demonstrates the utility of emerging data sources and methodologies for understanding playability on a much larger and more detailed scale than what has been done in previous playability research. Second, the theoretical contribution of this study is to explore playability at multiple spatial scales through the use of spatial regression models. To the best of our knowledge, no previous studies have examined playability using this approach at scale.

The remainder of this article is organized as follows. We provide a comprehensive literature review regarding playability-related research and using emerging data sources in auditing the urban environment in Section 2. In Section 3 we present the study area and the various data sources collected and used for conducting the experiments. In Section 4 we introduce the methods used, including crowd-sourced neighborhood playability labelling, advanced deep learning methods, and various regression approaches including ordinary least squares regression, geographically weighted regression, and multiscale geographically weighted regression. We highlight the results and discoveries in Section 4, showing the computed playability scores and associated factors. In Section 6 we discuss important findings and limitations of the study. Finally, in Section 7, we present our conclusions and vision for future research.

2. Literature review

Traditionally, studies examining the links between features of the built environment and active play in children have relied on a combination of two aspects: 1) surveys, in which children and/or parents respond with an estimation of play times and locations, and 2) the presence of urban physical attributes like open space and playgrounds (Aarts, de Vries, Van Oers, & Schuit, 2012; Laxer & Janssen, 2013). Similarly, some studies have used GPS trackers to monitor the duration and location of free play in children (Almanza, Jerrett, Dunton, Seto, & Pentz, 2012; Fjortoft, Lófman, & Halvorsen Thorén, 2010; Loebach & Gilliland, 2016). While survey methods can quantify how long and/or how frequently children play, such methods are costly and are thus generally limited to a particular time and place. Spatial representations of various urban features are useful for inference on play survey data, but typically only a handful of relevant features are available in such a format. Some researchers have also attempted to use satellite data as a



Fig. 1. A conceptual framework of playability.

new source for studying physical activity in children (Almanza, Jerrett, Dunton, Seto, & Pentz, 2012; Grigsby-Toussaint et al., 2011), but big data sources like this still only capture features of the physical environment (e.g. tree count and coverage). Finally, the amount of play-time spent outdoors by children has also been shown to decrease significantly with higher parent education (Aarts et al., 2012), suggesting that the playability of an environment can not necessarily be judged solely based on surveys of play-time and the presence or absence of physical features nearby.

Overall, these methods capture objective rather than subjective aspects of play. However, subjective experience of place is an important factor influencing what people do, and that includes play. For parents, perceived safety is an important determinant in allowing their children to play outside, while children might only want to play in attractive or fun-looking areas. Though some studies have used questionnaires to consider perception of safety with regards to play (Nguyen, Borghese, & Janssen, 2018), such studies have the same limitations regarding cost and extendability as other survey methods. To address this research gap, we propose to capture people's perception of the playability of neighborhoods by labelling street view imagery rather than by answering questionnaires, and then to extend the acquired knowledge to other areas by use of a (deep) machine learning model.

Recently, advances in big data, high-performance computing, and classification accuracy with deep neural networks have allowed researchers to study the urban environment at fine spatio-temporal scales, and in ways that capture elements of human perception not typically captured by traditional methods (Huang et al., 2020; Li, Zhang, & Li, 2015; Middel, Lukasczyk, Zakrzewski, Arnold, & Maciejewski, 2019; Yang, Huang, Li, Liu, & Hu, 2017; Yao et al., 2019; Yeh, Yue, Zhou, & Gao, 2020; Zhang et al., 2018). At the same time, street view images have arisen as a rich data source to comprehensively capture the environment from a human perspective. This data source, combined with the state-of-the-art deep learning models, allows researchers to describe entire cities at a sampling density of only a couple of meters by using the multiple view angles captured with street view images. Studies using this method have described various aspects of the urban physical environment, including place perception, housing prices, public health, and walkability (Geburu et al., 2017; Helbich et al., 2019; Hu et al., 2021; Kang et al., 2020; Li, Zhang, & Li, 2015; Zhang, Zhang, Liu, & Lin, 2018; Zhou, He, Cai, Wang, & Su, 2019). Furthermore, several similar studies have demonstrated how physical attributes of the built environment display substantial spatial heterogeneity, i.e. spatial variation in the feature, while other studies have shown spatial heterogeneity in perceptions of urban areas (Gao et al., 2017; Zhang, Fan, Kang, Hu, & Ratti, 2021; Zhang, Zhou, et al., 2018). Hence, we use this valuable data source to comprehensively audit the urban physical environment and relate it to the perception of playability, and then to examine associations between the derived playability scores and various geographical factors. Such an analysis allows a spatial understanding of playability at a city scale.

3. Data

The workflow of this paper is shown in Fig. 2, and has three components: data collection, scoring playability, and analysis between playability and geographic features. Two datasets are used in this research: 1) labelled street view imagery that represents the physical appearance and human perception of neighborhood playability, and 2) a multi-source dataset that includes a range of geographic features to represent the built environment, which enables us to examine associations between diverse urban features and neighborhood playability.

3.1. Study area

We take the cities of Boston, Seattle, and San Francisco as study areas, and data are collected in each city. These three cities are chosen as

they are large cities that represent different geographic areas of the contiguous United States, including both the East and West coasts. We aim to find general conclusions among different cities. Using these three cities also allows us to demonstrate the potential generalizability of the deep learning model.

All data layers are aggregated to the census block group (CBG) level. We use CBGs as the spatial analysis unit for two reasons: 1) all datasets are available or can be aggregated at the CBG level, and 2) the intended audience for this analysis is geographers, policy makers, city planners, and neighborhood residents, such that a finer spatial analysis might be too detailed.

3.2. Measuring the built-environment using street view images

Street view images that capture detailed scenery along road networks from a human perspective are collected in each city for scoring place playability. In total, 142,715 street view images for Boston, 255,293 images for Seattle, and 146,228 images for San Francisco are collected. With recent developments in computer vision and artificial intelligence techniques, street view images have become an effective data source for describing the real-world physical environment. Existing studies have already not only extracted elements and objects, but also understood human perceptions of streetscapes (e.g. lively, safe) from street view images (Helbich et al., 2019; Kang, Zhang, Peng, et al., 2020; Li, Zhang, & Li, 2015; Yao et al., 2019; Zhang, Zhou, et al., 2018). Such studies provide good foundations for both data sources and analytical techniques. Hence, street view images are used to comprehensively capture the built environment along streets as the basis for assigning playability scores.

To collect street view images, geo-referenced points are first generated every 100 m along the streets in three cities, Boston, Seattle, and San Francisco, respectively, using road networks downloaded from OpenStreetMap.² Utilizing the Google Street View API, view angles of 0, 90, 180, and 270 degrees are captured at each point, i.e. at least four images are downloaded for each geo-referenced point. Given that each point might have street view images from multiple years, only the most recent image for each location is used in our model training process.

3.3. Geographic features

A multi-source dataset of geographic features is constructed to analyze the relationships between playability and various urban features. In particular, we use 1) street speed limit and the number of crime incidents to represent neighborhood safety, 2) the number of trees and play areas nearby to represent the built environment of each neighborhood, and 3) human perceptions of the urban physical environment of a neighborhood. Attribute maps for each city can be seen in Figs. 3, 4, and 5, respectively.

The three different aspects of the urban environment listed above are modelled for the following reasons (which were also outlined in the conceptual framework of playability in Fig. 1). A place for children to play should be absolutely safe, with high traffic safety and a low crime rate. On the one hand, street speed limit might have a significant impact on determining if a place is suitable for children to play in. Road traffic crashes where an automobile hits a pedestrian are one of the leading causes of death for children worldwide (Stevenson, Sleet, & Ferguson, 2015). In such crashes, speed limit is an important factor since small increases in vehicle speed cause disproportionately large increases in mortality risk (Richards, 2010; W.H.O., 2015). Speed limit data, in km/h along a given segment of street, are acquired from the Boston Street Address Management system,³ Seattle GeoData,⁴ and the San Francisco

² <https://www.openstreetmap.org/>

³ <https://data.boston.gov/dataset/boston-street-segments>

⁴ <http://data-seattlecitygis.opendata.arcgis.com/datasets/seattle-streets>

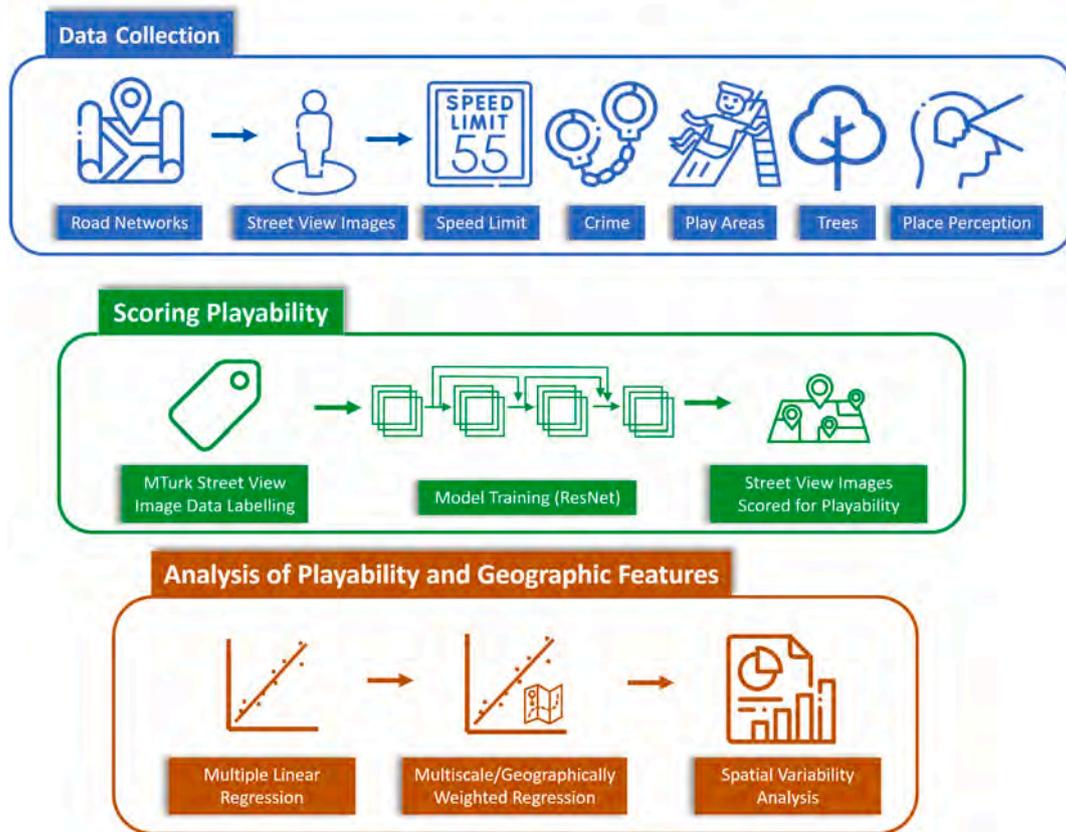


Fig. 2. Overview of our Workflow: (1) Data collection; (2) Scoring playability; (3) Analysis of playability and geographic features.

OpenData portal,⁵ respectively.

On the other hand, potential danger from threatening people is an important consideration for parents when determining if they should allow their children to play outside, and for children when looking for a place to play (Brussoni et al., 2020). Here, we model such risk as total crime incidents within 100-m of a given image location, using data from 2018 as provided by Boston Police Department,⁶ and the open data portals for the cities of Seattle⁷ and San Francisco,⁸ respectively.

We also calculate both play area count and tree count, as both are valuable factors that influence playability. Children need outdoor play spaces in which to play. To capture how such play areas might contribute to the playability of a place, we consider both parks and playgrounds available in a given location. Specifically, we count the number of such areas within a 200-m radius of the location of each street view image. Similarly, trees have been shown to be an affordance of play spaces, and have also been positively correlated with well-being in children (Kang, Zhang, Gao, Lin, & Liu, 2020; Kim, Lee, & Sohn, 2016).

Here, the number of trees along streets and in parks are counted within a 50-m radius of the location of each image, with only trees of significant size and under public management (i.e. city government) counted. Open space data were collected by searching for “park” and

“playground” in the open data portals for Boston,⁹ Seattle¹⁰ and San Francisco,¹¹ respectively. Tree data were collected from OpenTrees¹² for all cities.

In addition, we measure human perception of places by adopting the dataset and models described in Zhang, Zhou, et al. (2018). Human perception of places reflects the feelings people have when they experience (in this case, “see”) the urban physical environment, which here is captured by street view images. People have such perceptions when they consider the playability of places, as children are expected to play in places that are safe, comfortable, and fun. The *MIT Place Pulse*¹³ dataset captures various aspects of human perception with 110,988 labelled street view images from different locations across the globe, with labels provided by people from around the world. The results of the model trained on this dataset have been shown to be consistent and unbiased by demographics. We use this pre-trained model to evaluate our collected street view images for the *beautiful, boring, depressing, and lively* aspects of place perception in Boston, Seattle, and San Francisco.

4. Methods

As shown in Fig. 2, there are two parts of our methodology. The first part is scoring place playability. To do so, Amazon Mechanical Turk (MTurk) is employed for image labelling. Next, the images are used to train a deep neural network, which learns place playability from the labels and then scores all of the remaining street view images in three

⁵ <https://data.sfgov.org/Transportation/Speed-Limits/3t7b-gebh>

⁶ <https://data.boston.gov/group/public-safety>

⁷ <https://data.seattle.gov/Public-Safety/SPD-Crime-Data-2008-Present/tazs-3rd5>

⁸ <https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmnf-yvry>

⁹ <https://data.boston.gov/dataset/open-space>

¹⁰ <http://data-seattlecitygis.opendata.arcgis.com/datasets/seattle-parks>

¹¹ <https://data.sfgov.org/Geographic-Locations-and-Boundaries/San-Francisco-Open-Space-for-Shadow-Study-Analysis/xk8z-bcqz>

¹² <https://opentrees.org/>

¹³ <https://www.media.mit.edu/projects/place-pulse-new/overview/>

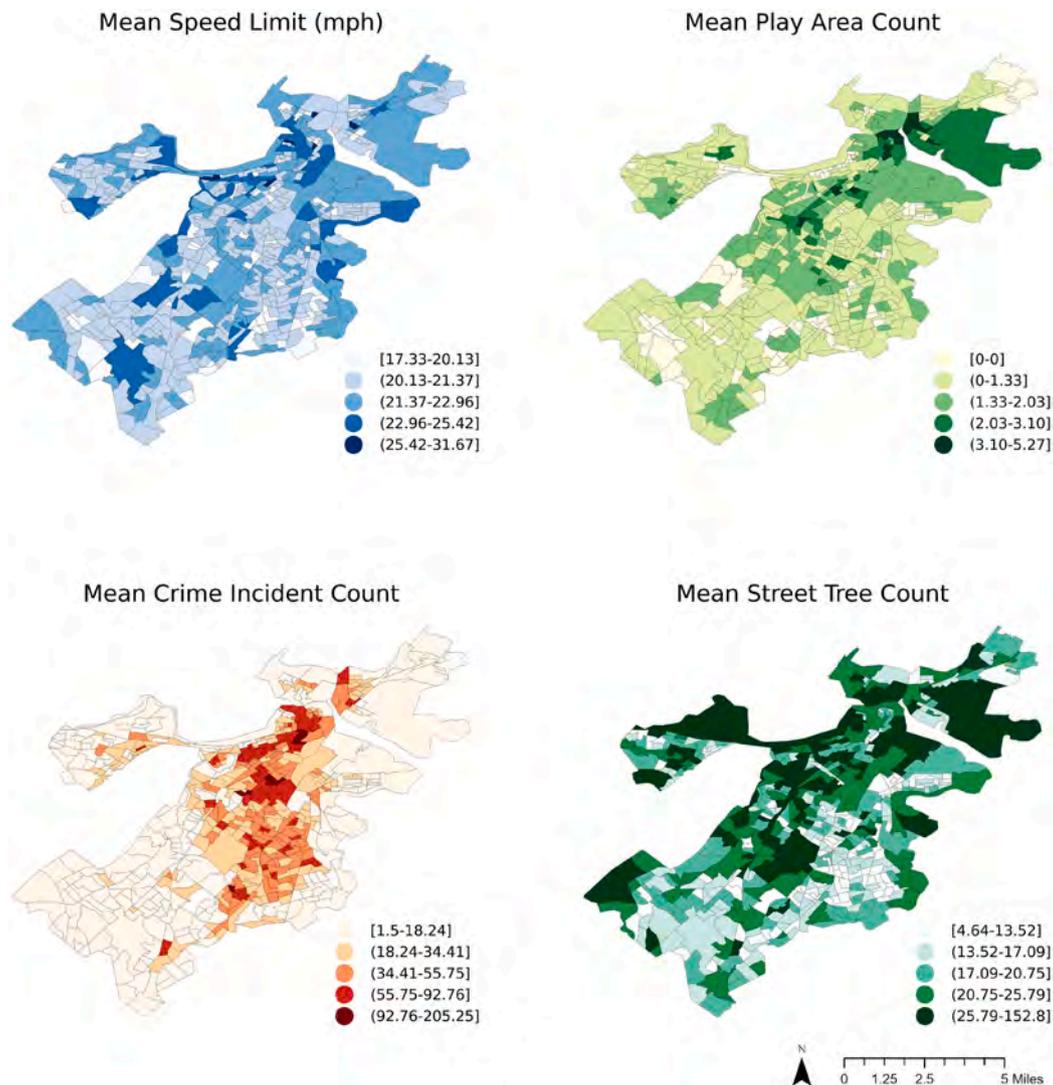


Fig. 3. The spatial distributions of the means of geographic features for Boston, including speed limit, play area count, crime count, and tree count.

cities to get a full description of neighborhood playability. The second part is analyzing associations between geographic features and playability. To do so, ordinary least squares (OLS) regression, geographically weighted regression (GWR), and multiscale GWR (MGWR) models are used to estimate the importance of several geographic variables on the playability values produced by the trained model. Most analyses are performed with the Microsoft Azure Cloud computing platform using Python 3.8 and the following statistical and machine learning packages: scikit-learn (Pedregosa et al., 2011), statsmodels (Seabold & Perktold, 2010), and mgwr-PySAL (Oshan, Li, Kang, Wolf, & Fotheringham, 2019). Additionally, the lctools package (Kalogirou, 2020) in R Studio version 4.1.0 is used to determine the optimal bandwidth for GWR.

4.1. Scoring playability

In this study, scoring the playability of places at a city scale is performed by asking volunteers to participate in a perception experiment, i. e. rating the playability of the scenery captured by street view images. To do so, the MTurk website is used for acquiring labelled playability data. MTurk is a crowd-sourcing marketplace that breaks down manual jobs, e.g. image classification labelling, into small tasks that can be

completed by workers from around the world.¹⁴ Recently, researchers have increasingly turned to MTurk for data labelling as it is fast, convenient, and low-cost compared to traditional survey methods; the accuracy is also good in place-based human intelligence tasks (Yan, Janowicz, Mai, & Gao, 2017). Given that human participants do the labelling in MTurk, approval from the University of Wisconsin Institutional Review Board was gained before the MTurk labelling was initiated.

For this study, workers on MTurk are asked: “Choose the category that most accurately describes how playable the scene in the image is for children.” By doing so, workers are able to consider how playable the scene in a street view image would be for children, and then choose the appropriate label from among five choices: *very unplayable*, *unplayable*, *neither playable nor unplayable*, *playable*, *very playable*. The web interface for workers to label street view images is shown in Fig. 6. In total, 3011 street view images are randomly picked from the entire collection of Boston street view images and used for collecting workers’ perceptions. Playability is not defined in the MTurk portion of the study because we want to get people’s general understanding of neighborhood playability, rather than relying a precise, technical definition. Such an approach allows for different individual and cultural understandings of playable

¹⁴ <https://www.mturk.com/>

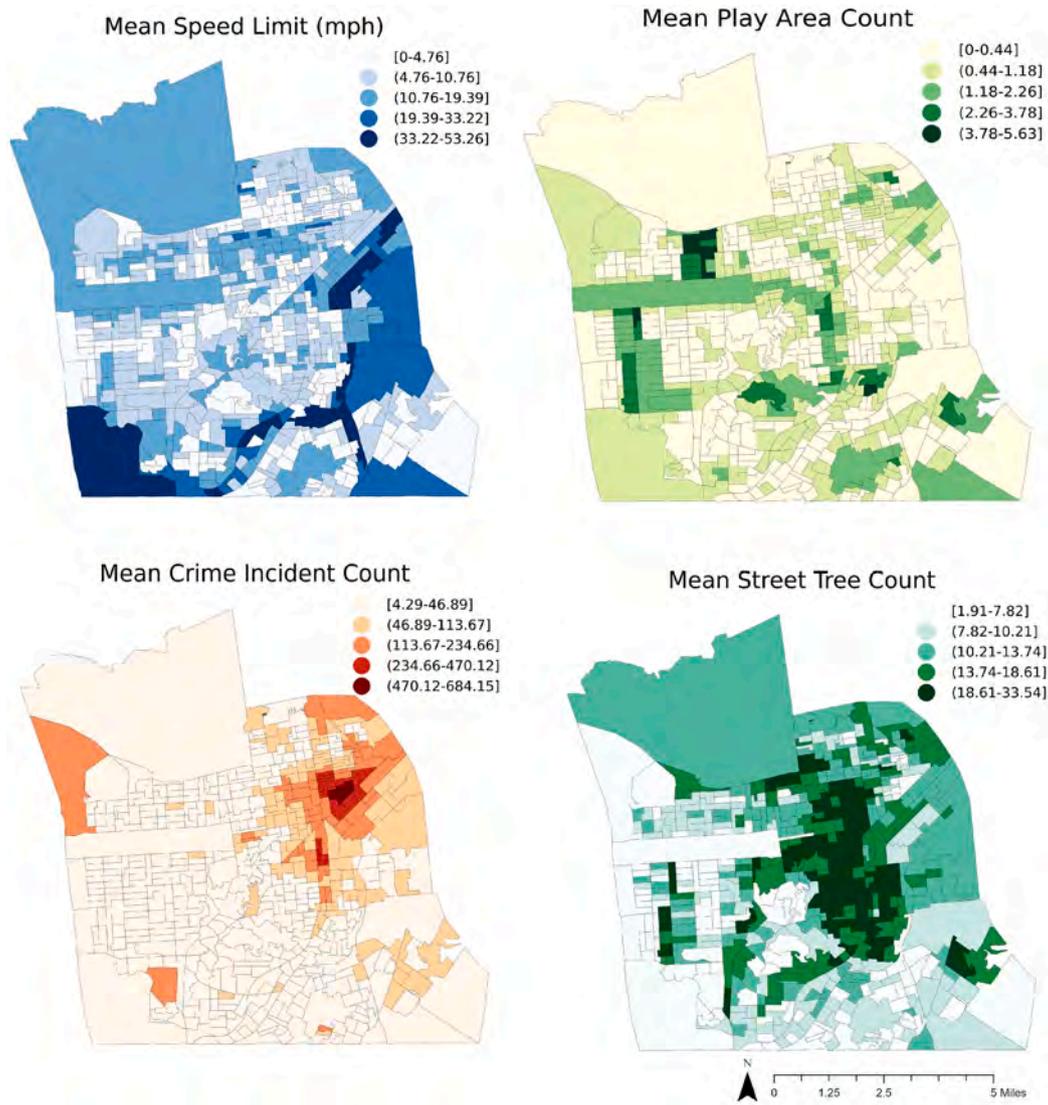


Fig. 4. The spatial distributions of the means of geographic features for San Francisco, including speed limit, play area count, crime count, and tree count.

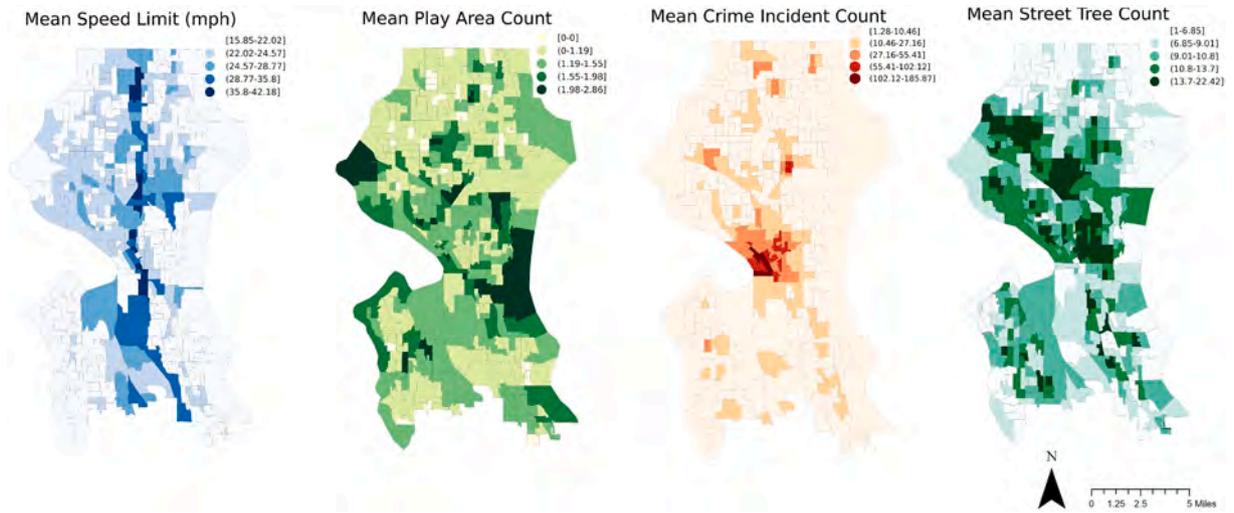


Fig. 5. The spatial distributions of the means of geographic features for Seattle, including speed limit, play area count, crime count, and tree count.



Fig. 6. Amazon Mechanical Turk (MTurk) User Interface for Classification.

places, which is important given that there can be significant variation within and between groups of people regarding their perception of places (Gao et al., 2017; Yao et al., 2019). To arrive at a consensus understanding of playability, each of the images collected is labelled by nine workers, with a total of 210 workers contributing to the labelling of the entire training set. We assign scores for different labels ranging from *very unplayable* (score = 1) to *very playable* (score = 5), and the average of the nine labels for each image is used as the playability score.

4.2. Deep learning model

After collecting the playability scores of the selected 3,011 street view images described above, a pre-trained ResNet18 deep neural network model (He, Zhang, Ren, & Sun, 2016) is used to learn characteristics of playability from the labelled street view images and then to score all of the unlabelled street view images. The training process was formulated as a five-category classification task, i.e. to classify each image as *very unplayable*, *unplayable*, *neither playable nor unplayable*, *playable*, or *very playable*. High dimensional visual features that contain meaningful semantics for representing neighborhood playability are extracted by the deep neural network. Cuda and PyTorch in Python 3.8 are used to perform the training and evaluate the model. Once the model is fully trained, it is used to evaluate the collected thousands of street view imagery dataset for Boston, San Francisco, and Seattle, the scores of which are then used as the predicted playability values in regression analysis. After calculating the playability scores for all of the street view images, the average playability scores are aggregated to the CBG level. An average value of playability can potentially reduce spatial non-stationarity and the standard deviation of scores to derive the common perception trend of a neighborhood.

4.3. Regression

After computing the place playability scores, we examine the associations between playability and collected geographic features in cities. To understand where and how these geographic features affect neighborhood playability, three different regression models are used including ordinary least squares (OLS) linear regression, geographically weighted regression (GWR), and multiscale GWR (MGWR).

Linear regression using ordinary least squares (OLS) as a parameter estimation method has been widely used to examine the associations between dependent and independent variables. OLS linear regression is a global model that estimates unknown parameters (a_k) in a linear model by minimizing the sum of squared differences between predicted values and observed values in the dataset. The equation is as follows:

$$Y_i = \alpha_0 + \sum_{k=1}^m a_k X_k + \varepsilon_i \quad (1)$$

With the OLS linear regression model, spatial stationarity is assumed,

meaning that neither the distance between data points nor the location of data points is assumed to impact the values of those data points. As it includes all data points in parameter estimation, the resulting coefficients (including a_k and the intercept α_0) are global averages. Many phenomena, however, do vary across space, i.e. the effects might be distinct at different places, and thus are better modelled by spatial models. A popular model that takes spatial heterogeneity into consideration is geographically weighted regression (GWR). GWR is a spatial regression technique that takes spatial non-stationarity into account by building a number of local regression models using the following equation from Fotheringham, Brunson, and Charlton (2003):

$$Y_i = \alpha_{0(u_i, v_i)} + \sum_{k=1}^m a_{k(u_i, v_i)} X_{k(u_i, v_i)} + \varepsilon_i \quad (2)$$

where Y_i refers to the playability score i with its coordinates (u_i, v_i) ; $\alpha_{0(u_i, v_i)}$ denotes the intercept; $a_{k(u_i, v_i)}$ indicates the local regression coefficient for the k^{th} independent variable at location (u_i, v_i) ; $X_{k(u_i, v_i)}$ refers to the k^{th} attribute of location i ; and ε_i is the random error. By using this model, the derived coefficients can vary across different sub-areas, and we are able to understand how the effects of geographical features on playability change across space.

Since GWR uses a number of local models, the neighborhood, i.e. what points will be included in each model, must be determined. With the assumption that all of the parameters being estimated vary on the same spatial scale, a single bandwidth is used to define the neighborhood. To allow for interpretability, the distance-based bandwidth is presented in terms of meters. By using a bi-square kernel, all points within the bandwidth distance are inversely weighted with the distance from the center of the local area, and points outside of the bandwidth are assigned a weight of zero. The optimal bandwidth is chosen via cross-validation with the `lctools` package in R Studio (Kalogirou, 2020), while all other regression-related tasks are performed in Python with the `mgwr` package (Oshan et al., 2019).

For model evaluation, we use a golden section search optimization routine with a corrected Akaike information criterion (AICc) as the model criterion. For interpretation of the model, local R^2 and global AICc are considered for goodness-of-fit. Finally, the parameter estimates for the predictors at each location are then mapped across the region of study, showing how parameters values vary across space.

While GWR can improve goodness-of-fit when modelling spatial processes, it does so by applying a universal bandwidth to all of the parameters being estimated. While the spatial scale of all considered variables may be the same in some cases, there are many cases in which the processes generating the data may differ from variable to variable. To deal with this problem, multiscale (M)GWR allows the bandwidth to vary locally for each variable, such that the bandwidth for each variable is adjusted to the spatial scale of that variable. Allowing variables to use different spatial scales reduces bias in the parameter estimates, over-

fitting, and concurvity. The equation from Fotheringham, Yang, and Kang (2017) is as follows:

$$Y_i = \alpha_{0(u_i, v_i)} + \sum_{k=1}^m a_{bwk(u_i, v_i)} X_{k(u_i, v_i)} + \epsilon_i \quad (3)$$

The equation is adapted from the GWR equation by the addition of *bw* bandwidth term for the k^{th} attribute. Different from GWR, we use an adaptive, nearest-neighbor bandwidth kernel, which specifies the number of neighbors that must be included in each local regression model. Such an approach better handles edge effects and non-uniform spatial distributions (Oshan et al., 2019), which may be more prominent given that each variable has its own spatial scale in MGWR. Like with GWR, a bisquare distance-weighting function is used. For parameter optimization, we use a golden section search optimization routine for parameter optimization, with AICc as the model goodness-of-fit criterion. R^2 and AICc are used for the model fit evaluation, and parameter estimates are mapped as with GWR.

5. Results

5.1. Playability score

We first introduce the results of playability labelling by MTurk (in Fig. 7). The average image labels supplied by MTurk are mostly concentrated in the range of *very unplayable* through *playable*, with the *playable* category having the highest number of images and *very playable* having the fewest. For each category, the percentage of total training data is as follows: 23% *very unplayable*, 20% *unplayable*, 20% *neither playable nor unplayable*, 26% *playable*, and 12% *very playable*. For the model output category distribution, the percentage of each class is as follows: 23% *very unplayable*, 21% *unplayable*, 13% *neither playable nor unplayable*, 37% *playable*, and 6% *very playable*. As seen in Fig. 7, the distributions of the training data and the model output for Boston are very similar, suggesting that the sampling method effectively captured the Boston street view imagery, and that the deep learning model effectively learned playability features. Since the deep learning model indeed learns characteristics that represent the playability of a neighborhood, it can be used to score the playability of scenes captured in street view images.

To quantify the reliability of our results and the ability of the model to generalize, we present here the classification accuracy for each city. In

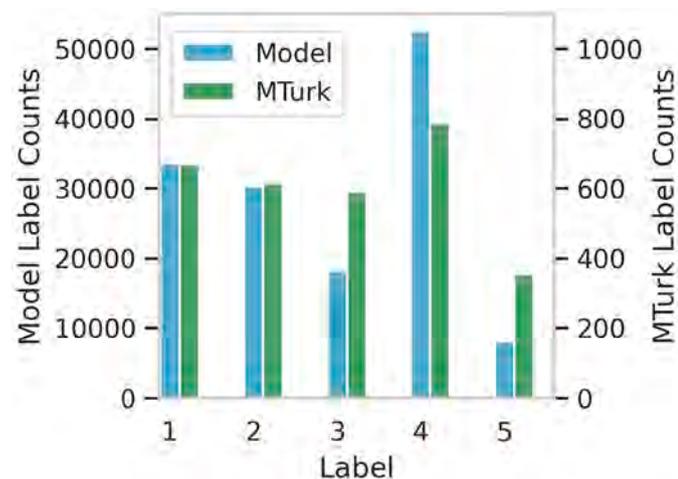


Fig. 7. Distributions of playability labels. Green: labels from Amazon Mechanical Turk (MTurk); Cyan: labels from the trained Resnet18 model. On the x-axis: 1, 2, 3, 4, and 5 correspond to *very unplayable*, *unplayable*, *neither playable nor unplayable*, *playable*, and *very playable*, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

each city, the top-2 classification accuracy is used, such that if the ground-truth label is either the highest or second highest probability label as predicted by the trained ResNet model, a value of 1 is assigned to the image, and a value of 0 if otherwise. The mean accuracy is then taken over all images with ground truth data to produce classification accuracy. We consider top-2 accuracy to better capture model performance compared to top-1 accuracy, since the playability labels are ordinal and people have variability in their perceptions, especially for the two similar categories such as “playable” and “very playable”. Top-2 accuracy for Boston is produced from the validation set for the training epoch associated with the highest-performing model, since the model was trained on labelled street view images from Boston, while a separate MTurk data-labelling workflow following that described in the Methods section is used to provide ground-truth labels for Seattle and San Francisco. The resulting top-2 classification accuracy for Boston, Seattle, and San Francisco is 79%, 70%, and 65%, respectively. These results indicate that the methods used can generalize fairly well between the cities studied, though some accuracy is lost when evaluating images from cities not represented in the training set. The fact that the classification accuracy is so high for the cities not represented in the training data speaks to the similarities in human conceptions of playability even across different cities. To verify that the training sample size is sufficiently large, we also test the top-2 accuracy in Boston for the model trained on a subset of only 2000 images from the Boston image data set. The resulting top-2 accuracy for Boston is 74%, compared to 79% with the full training set, suggesting that additional training samples over 3000 would only marginally improve the model accuracy.

To demonstrate the similarity between human-assigned labels and model-assigned labels, we present Fig. 8, which shows some samples of street view images that are classified into different categories. Street view images which are classified as *very unplayable* feature large roads and little green space, while images classified as *very playable* are primarily from suburbs with trees, narrow roads, houses, and sidewalks. As expected, images labelled *neither playable nor unplayable* have a mix of features from both *very unplayable* and *very playable* images, and the features present are not as pronounced: roads are medium-sized, yards are present but small, and relatively few trees are close by. Analysis with explanatory variables is described in the section 5.2.

The distributions of average playability values per CBG in three cities are shown in Fig. 9. According to the maps, Boston playability values are highest in the southern part of the city, and lowest near the downtown and airport areas in the northern part of the city. In San Francisco, scores are highest in the geometric center of the city and in the large park on the northern shore, while being relatively low in the rest of the city. In Seattle, the most playable areas are in the north and on the perimeter of the city.

5.2. Regression results

Six experiments are performed using three models (OLS, GWR, and MGWR) to examine associations between neighborhood playability and different geographic features in three cities, respectively:

- 1) M1: OLS model with only speed limit, crime rate, tree count, and play area count.
- 2) M2: OLS model with all variables, including place perception variables.
- 3) M3: GWR model with only speed limit, crime rate, tree count, and play area count.
- 4) M4: GWR model with all variables, including place perception variables.
- 5) M5: MGWR model with only speed limit, crime rate, tree count, and play area count.
- 6) M6: MGWR model with all variables, including place perception variables.



Fig. 8. Samples of labelled street view images classified by the deep learning model. Left: street view images that represent very unplayable environments; Middle: street view images that represent neither unplayable nor playable environments; Right: street view images that represent very playable environments.

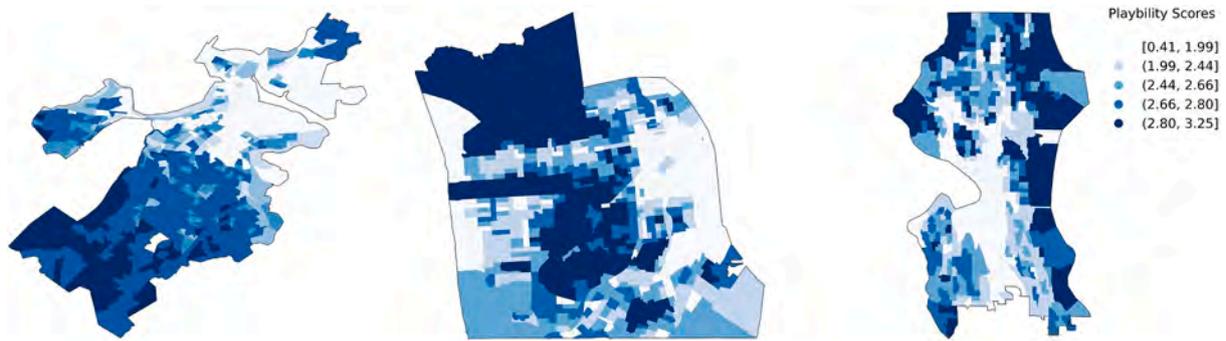


Fig. 9. Playability scores from the model prediction at the census block group scale for Boston (Left), San Francisco (Center), and Seattle (Right).

Using cross-validation, the optimal bandwidths for the GWR models in each city are as follows. For Boston GWR models without and with place perception variables: 14201.76 m and 13,506.89 m, respectively. For San Francisco GWR models without and with place perception variables: 4489.18 m and 4337.99 m, respectively. For Seattle GWR models without and with place perception variables: 10175.22 m and 16,301.80 m, respectively. The optimal GWR bandwidths for Seattle and Boston have similar spatial scales, while that of San Francisco is substantially smaller. This suggests that variables considered in Seattle and Boston operate on a larger spatial scale than the same variables in San Francisco, which might link to their underlying urban built

environments.

Overall, playability values are best modelled with the full set of variables and with models that account for spatially-heterogeneous processes (Tables 1, 2, 3). Model goodness-of-fit (R^2) improves significantly when taking spatial heterogeneity into consideration with GWR and MGWR, compared to OLS regression.

The inclusion of place perception variables improves the R^2 as well, which suggests that modelling with place perception variables captures more environmental factors that are associated with playability. In Boston, for instance, the R^2 increases from 0.33 to 0.63 for the OLS model with the inclusion of the place perception variables. Regarding

Table 1
Boston regression results.

	OLS Coefficients		GWR Coefficients				MGWR Coefficients			
	M1	M2	M3	M4		M5	M6			
Variables	Mean	Mean	Mean	Min, Max	Mean	Min, Max	Mean	Min, Max	Mean	Min, Max
Intercept	0.00	0.00	-0.08	(-0.96, 0.70)	-0.05	(-0.74, 0.86)	-0.04	(-1.44, 0.86)	-0.07	(-0.92, 0.72)
Speed Limit	-0.34***	-0.25***	-0.26	(-0.60, 0.05)	-0.17	(-0.39, 0.02)	-0.22	(-0.40, -0.08)	-0.14	(-0.24, -0.10)
Crime	-0.35***	-0.11***	-0.35	(-1.31, 0.22)	-0.26	(-1.14, 0.26)	-0.29	(-0.63, -0.11)	-0.24	(-0.72, 0.02)
Tree Count	-0.06	-0.04	-0.02	(-0.27, 0.16)	-0.02	(-0.11, 0.28)	-0.02	(-0.03, -0.01)	-0.03	(-0.03, -0.03)
Play Area Count	-0.17***	-0.09**	-0.02	(-0.50, 0.20)	-0.02	(-0.43, 0.16)	0.01	(-0.07, 0.06)	0.00	(-0.12, 0.22)
Beautiful		0.37***			0.16	(-0.18, 0.44)			0.13	(0.02, 0.25)
Boring		-0.23***			-0.16	(-0.50, 0.41)			-0.10	(-0.13, -0.09)
Depressing		-0.28***			-0.24	(-0.42, 0.06)			-0.26	(-0.26, -0.25)
Lively		-0.46***			-0.18	(-0.59, 0.38)			-0.01	(-0.21, 0.03)
Observations	558	558	558		558		558		558	
R^2	0.33	0.63	0.76		0.86		0.79		0.88	
AICc	1371.47	1048.42	885.14		664.53		817.67		571.90	

R^2 and Adj R^2 were nearly identical, and so only R^2 is reported.

*** denotes significance at the 0.1% level, ** at the 1%, and * level 5% level.

Table 2
San Francisco regression results.

Variables	OLS Coefficients		GWR Coefficients		MGWR Coefficients					
	M1	M2	M3	M4	M5	M6				
Intercept	Mean 0.00	Mean 0.00	Mean -0.02	Min, Max (-0.68, 0.43)	Mean -0.07	Min, Max (-0.43, 0.35)	Mean -0.26	Min, Max (-1.00, 0.38)	Mean -0.21	Min, Max (-0.63, 0.07)
Speed Limit	-0.28***	-0.36***	-0.30	(-0.51, 0.09)	-0.39	(-0.56, -0.11)	-0.33	(-0.97, 0.23)	-0.39	(-0.98, -0.04)
Crime	-0.46***	-0.26***	-0.52	(-1.71, 0.91)	-0.53	(-1.3, -0.10)	-0.79	(-2.30, 0.35)	-0.67	(-2.16, 0.16)
Tree Count	0.38***	0.29***	0.37	(0.10, 0.59)	0.25	(-0.17, 0.46)	0.32	(-0.06, 0.76)	0.24	(0.10, 0.37)
Play Area Count	0.11***	0.06*	0.13	(-0.24, 0.57)	0.10	(-0.11, 0.53)	0.21	(-0.48, 1.01)	0.13	(-0.16, 0.57)
Beautiful		0.30***			0.23	(-0.12, 0.70)			0.21	(0.01, 0.47)
Boring		0.06			0.09	(-0.40, 0.23)			0.11	(0.10, 0.13)
Depressing		-0.35***			-0.37	(-0.70, 0.06)			-0.30	(-0.31, -0.29)
Lively		-0.44***			-0.31	(-0.93, 0.25)			-0.16	(-0.57, 0.15)
Observations	563	563	563		563		563		563	
R ²	0.46	0.65	0.67		0.78		0.84		0.87	
AICc	1260.95	1022.95	1035.57		859.18		831.20		729.74	

R² and Adj R² were nearly identical, and so only R² is reported.
*** denotes significance at the 0.1% level, ** at the 1% level, and * level 5% level.

Table 3
Seattle regression results.

Variables	OLS Coefficients		GWR Coefficients		MGWR Coefficients					
	M1	M2	M3	M4	M5	M6				
Intercept	Mean 0.00	Mean 0.00	Mean -0.22	Min, Max (-4.83, 1.20)	Mean -0.12	Min, Max (-0.60, 0.25)	Mean -0.28	Min, Max (-1.3, 0.36)	Mean -0.03	Min, Max (-0.37, 0.29)
Speed Limit	-0.46***	-0.31***	-0.46	(-1.50, 1.02)	-0.33	(-0.72, -0.12)	-0.46	(-0.91, -0.19)	-0.32	(-0.66, -0.11)
Crime	-0.59***	-0.23***	-1.00	(-9.44, 1.17)	-0.47	(-1.83, 0.01)	-0.94	(-1.84, -0.26)	-0.36	(-0.69, -0.10)
Tree Count	0.02	0.00	0.13	(-0.89, 1.71)	0.01	(-0.23, 0.26)	0.16	(-0.25, 0.97)	0.03	(-0.02, 0.07)
Play Area Count	-0.03	-0.03	0.04	(-0.31, 0.69)	0.00	(-0.14, 0.24)	0.05	(-0.03, 0.16)	0.02	(0.01, 0.03)
Beautiful		0.43***			0.39	(0.10, 0.82)			0.33	(0.07, 0.73)
Boring		0.07			0.03	(-0.35, 0.40)			-0.04	(-0.14, 0.04)
Depressing		-0.27***			-0.27	(-0.69, 0.32)			-0.33	(-0.48, -0.19)
Lively		-0.17***			-0.19	(-0.47, 0.52)			-0.29	(-0.43, -0.10)
Observations	479	479	479		479		479		479	
R ²	0.62	0.814	0.86		0.91		0.85		0.93	
AICc	906.07	573.10	672.96		426.01		663.06		311.83	

R² and Adj R² were nearly identical, and so only R² is reported.
*** denotes significance at the 0.1% level, ** at the 1% level, and * level 5% level.

the two spatial models, the goodness-of-fit improved from 0.76 to 0.86 and from 0.79 to 0.88 for GWR and MGWR, by without considering or with considering the place perception variables, respectively. Regression results are similar for both San Francisco (Table 2) and Seattle (Table 3), suggesting that similar factors are associated with playability in different parts of the United States, at least in our selected cities of comparable size.

Using the AICc (Hurvich & Tsai, 1993) as an estimator of prediction error is better suited to handle small sample sizes and account for model complexity; the smaller the AICc value the better the model. The MGWR for Boston yields a value of 571.90, compared to 664.53 from GWR (Table 1), when place perception variables are included, suggesting that

MGWR does produce a better fit when compared with GWR, while not significantly increasing model complexity. Similar AICc values are produced from GWR and MGWR models in Seattle and San Francisco. However, given the high goodness-of-fit from the inclusion of the place perception variables and the relative ease of interpretation of GWR compared to MGWR, further discussion will only consider the GWR model constructed with the place perception variables.

Local R² results for the GWR model constructed for Boston show that highest model goodness of fit is achieved near the downtown area, though the R² is relatively high throughout the city (Fig. 10). Seattle and San Francisco, similarly, have high accuracy over the whole area of study, though with some spatial variation.

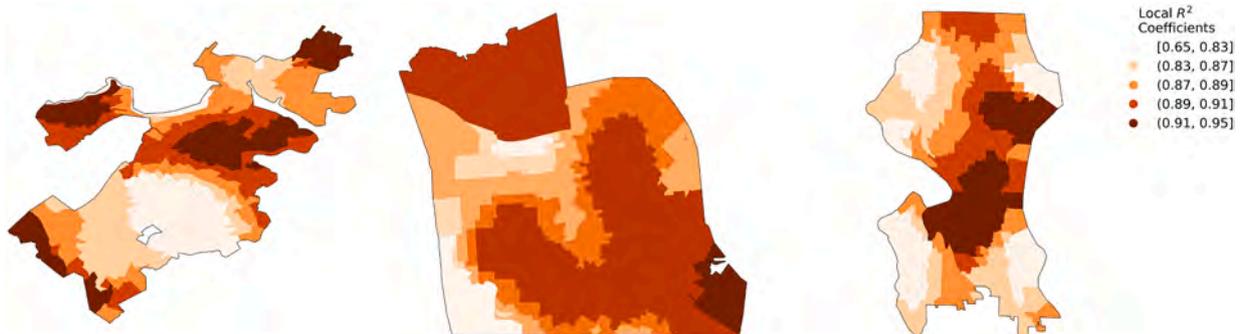


Fig. 10. Local R² for GWR model in Boston (Left), San Francisco (Center), and Seattle (Right).

Here we present the model coefficients for the GWR models with place perception variables. Within Boston, the place perception variables are the most predictive variables, with average coefficient values of 0.16, -0.16, -0.24, and -0.18 for the beautiful, boring, depressing, and lively variables, respectively (Table 1). Intuitively, beautiful is positively correlated with playability, while boring and depressing are negatively correlated with playability. In San Francisco (Table 2) and Seattle (Table 3), coefficient values for beautiful, depressing, and lively are all similar to those of Boston, while boring has small but positive average coefficient values in both cities.

While it is intuitive that boring or depressing would be negatively associated with playable places, the negative relationship between lively and playable is more surprising. Liveliness is generally considered to be a positive attribute of place, so here we explore where and why liveliness and playability scores are negatively correlated. The negative relation between lively and playable suggests that the activities, and thus places, which are associated with liveliness are adult activities, not child activities, and vice-versa. To visualize where differences between playability and liveliness are greatest, we perform the following operations: 1) normalize scores for liveliness and playability to be between 0 and 1, and 2) take the difference between the two attributes by subtracting normalized playability scores from normalized liveliness scores at the CBG level (Fig. 11). Values closer to -1 have relatively low liveliness and relatively high playability while values closer to +1 have relatively high liveliness and relatively low playability. In particular, we find that downtown areas in the three cities all have relatively high liveliness scores and relatively low playability scores. As seen in Fig. 11, the most negative areas in Boston are in the southern, more residential part of the city, indicating relatively low liveliness and high playability. The opposite is true for the northern half of the city, which is generally more commercial and industrial. In San Francisco, low liveliness and high playability are seen in the large park on the north side (in white color). In Seattle, commercial and industrial areas (in downtown) generally have high liveliness and low playability (in purple color), but other areas in the city show no clear patterns following predominate land-use patterns, similar to San Francisco.

Speed limit in all three cities has a strong, negative correlation with playability (see Fig. 12), with average coefficient values of -0.17, -0.39, and -0.33 in Boston, San Francisco, and Seattle. Crime, similarly, has a strongly negative correlation (-0.26, -0.53, -0.47, respectively) with playability in all cities.

Average coefficient values for tree count within 50 m range from positive to negative in their association with playability (and the spatial variability is shown in Fig. 13), with average coefficients of -0.02, 0.25, and 0.01 for Boston, San Francisco, and Seattle. Generally, positive coefficients for tree count are located on the periphery of each of the cities, perhaps coinciding with more suburban neighborhoods and parks, though that is not the case in southern Boston and southern Seattle, which both have many parks and suburbs yet still have negative coefficients. It demonstrates the spatial heterogeneous impact of tree count

on playability.

GWR coefficient values for play area count are generally close to zero (Fig. 14), varying between slightly positive and slightly negative. Areas with positive associations between play area count and playability are located along the periphery of the three cities. It also shows the spatial heterogeneous impact of play area count on playability.

5.3. Spatial variability testing results

To assess if the models are truly capturing spatial heterogeneity, a Monte Carlo simulation-based statistical significance test is performed wherein the locations of the attributes are shuffled a specified number of times and regression models are built on each set of shuffled data, allowing for the construction of a pseudo *p*-value for each parameter estimate distribution. Using a 95% confidence level, a *p*-value of 0.05 or less for a given parameter indicates that the variable has a significant level of spatial variability, i.e. a non-random spatial distribution, while values greater than 0.05 indicate that the spatial arrangement is possibly random.

Due to the intense computational requirements for such a test, the spatial variability test for GWR on each set of variables is performed with 200 iterations, while MGWR is performed with 100 iterations on each set of variables. Results for the spatial variability tests with the GWR models built with and without the place perception variables indicate statistically significant spatial variability, i.e. non-randomness, with all variables having a *p*-value of ≤ 0.05 (Table 4). For MGWR without place perception variables, tree count and play space count are not significant in Boston, while all other variables are. Using MGWR with all variables, the following variables have significant spatial variability at *p*-value ≤ 0.05 : 1) in Boston, crime, beautiful, and lively; 2) in San Francisco, speed limit, crime, play area count, beautiful, and lively; 3) in Seattle, speed limit, crime, beautiful, depressing, and lively. These results suggest that with more variables and different bandwidths for each variable, the interaction between each predictor and the dependent variable becomes smaller relative to the variation in value across the study area, resulting in insignificant spatial variability for some variables.

6. Discussion

Here we summarize key findings from the regression analyses of playability in the three cities, then discuss limitations of the study.

6.1. Factors related to playability

The results described above are generally consistent both with previous studies and common sense. For instance, our results suggest that places that are perceived as beautiful are generally perceived as playable. However, previous research has identified trees as being associated with more play (Dyment & Bell, 2008), while our results suggest that the

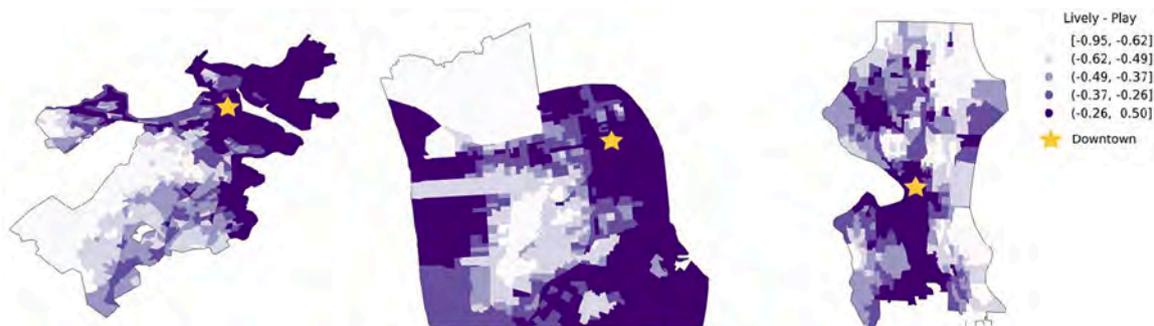


Fig. 11. Lively and playability differences in Boston (Left), San Francisco (Center), and Seattle (Right). Golden stars are placed in the downtown commercial districts to give an indication of the respective city centers, rather than to comprehensively describe cognitive place.

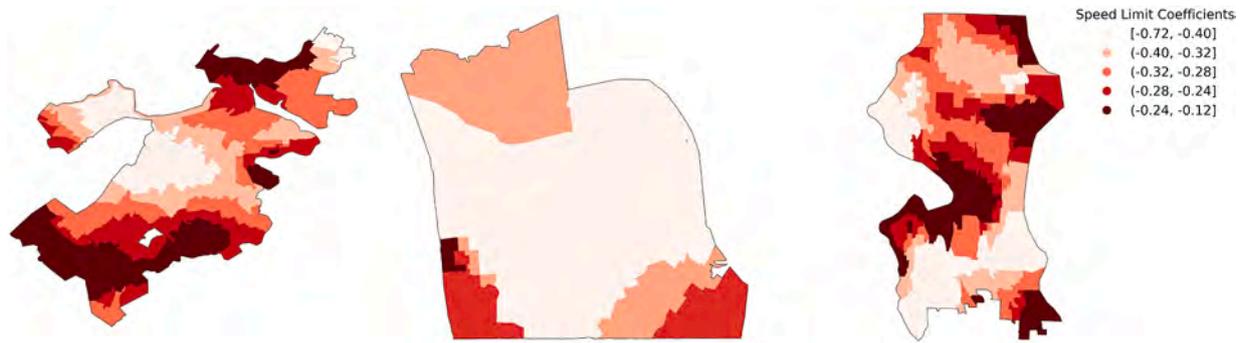


Fig. 12. Speed limit GWR coefficients for Boston (Left), San Francisco (Center), and Seattle (Right).



Fig. 13. Tree count GWR coefficients for Boston (Left), San Francisco (Center), and Seattle (Right).

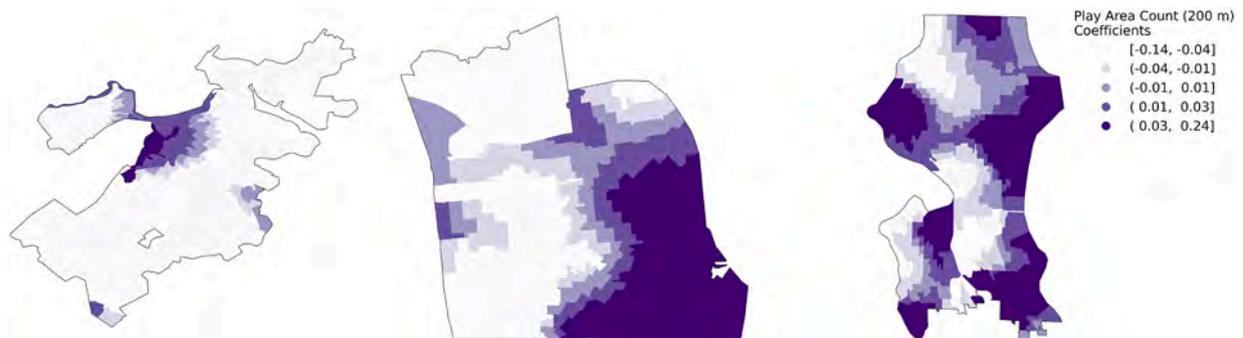


Fig. 14. Play area count GWR coefficients for Boston (Left), San Francisco (Center), and Seattle (Right).

relationship between tree count and playability varies geographically, though this could be a result of only including publicly-managed trees along roads and in parks in our tree count dataset.

While previous studies have found different conclusions regarding traffic calming features and vehicle speed (Lambert, Vlaar, Herrington, & Brussoni, 2019; Nguyen et al., 2018), we find here that higher road speeds have a strongly negative association with playability in all places. Higher road speed may affect playability in several ways: 1) increase the risk of a fatal collision between an automobile and a child, such that parents would prevent children from playing for safety reasons, 2) increase the noise level from traffic, thereby making play in such areas undesirable, and 3) indirectly, roads with higher speed limits may be associated with smaller or non-existent sidewalks rendering them less playable.

Crime count also has a strong, negative association with playability and may affect playability in several ways. Since crime count here is merely the sum total number of crimes in the vicinity of each street view image, the type of crime is obscured. For example, petty crimes or crimes that occur at night might not affect playability, while crimes that

directly affect the safety of children might have a stronger impact. For both road speed and crime, targeted approaches to improve playability in a given neighborhood would need to consider such distinctions.

Generally, the play area count data set used here has a weak relationship with playability. Common sense, however, says that access to play areas is essential in being able to play. This discrepancy may be due to the lack of some informal places, such as cul-de-sacs and empty lots, in our data set, as data is downloaded from the VGI platform OSM. More data sources may be obtained to have a more comprehensive description of place.

Regarding place perception factors, higher place perception attribute scores for depressing and lively are both negatively associated with playability, while boring varies from negative to positive in its association with playability.

6.2. Differences between playability and liveliness

We continue here the discussion based on the differences between lively and playable scores, as initially described in Section 5.2.

Table 4
Spatial variability test results.

Variables	Boston		San Francisco		Seattle	
	GWR	MGWR	GWR	MGWR	GWR	MGWR
	<i>p-value</i>					
Intercept	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)
Speed Limit	(0.00, 0.00)	(0.05, 0.12)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)
Crime	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)
Tree Count	(0.00, 0.00)	(0.85, 1.00)	(0.00, 0.00)	(0.00, 0.06)	(0.00, 0.00)	(0.00, 0.17)
Play Area Count	(0.01, 0.00)	(0.39, 0.06)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.54)
Beautiful	(-, 0.00)	(-, 0.05)	(-, 0.01)	(-, 0.01)	(-, 0.01)	(-, 0.02)
Boring	(-, 0.00)	(-, 0.57)	(-, 0.01)	(-, 0.37)	(-, 0.00)	(-, 0.09)
Depressing	(-, 0.00)	(-, 0.87)	(-, 0.00)	(-, 0.80)	(-, 0.01)	(-, 0.05)
Lively	(-, 0.00)	(-, 0.05)	(-, 0.00)	(-, 0.00)	(-, 0.00)	(-, 0.04)

In each city, spatial variability tests are performed on both GWR and MGWR models. Additionally, each model type was constructed both with and without the place perception variables.

Referencing the land use maps in three cities^{15 16 17}, we discuss reasons that might cause such differences, and explore common themes. Generally, the difference maps suggest that liveliness and playability both stem from land-use policies but are generally impacted in opposite ways. In all three cities, there are clear geographic patterns in the relative strength of each score, but the patterns vary by city. However, one consistent finding across all three cities is high liveliness and low playability in the downtown area, which we mark with a golden star in each city in Fig. 11. Stars are placed in the downtown commercial districts to give an indication of the respective city centers, rather than to comprehensively describe cognitive places (Gao et al., 2017; Montello, Goodchild, Gottsegen, & Fohl, 2003).

While lively downtowns may be desirable for affluent young adults and tourists, they do not necessarily cater to all people. Studies by Gilliland, Holmes, Irwin, and Tucker (2006) and Whitzman and Mizrachi (2012) found low access to play spaces and a sense of entrapment by children in downtown areas. A study by Quercia, O'Hare, and Cramer (2014) that used crowd-sourcing and image-processing techniques to quantitatively understand the built environment found that tall residential buildings are negatively associated with beauty and happiness, while tall office buildings and landmarks are positively associated with beauty, while also providing a visual cue of city structure. In the cities examined in this study, the downtown areas are composed of many tall office buildings, suggesting that they may provide beauty and orientation cues, which could be useful to adults using the downtown for work and night-life activities, but fail to create places that generate a sense of playability. For another aspect of urban vibrancy, a study by Zhou et al. (2019) found that walkability scores produced with a street view image and deep learning method did not have a strong relationship with urban density. Finally, De Nadai et al. (2016) found that the perceived safety of an area modulates the human activity there, with females and people older than 50 being particularly sensitive to the level of perceived safety. While the same study does not examine the effects of perceived safety on the behavior of children, we note that our results show a strongly negative relationship between playability and crime, with crime counts

being highest in all cities near the downtown area (Figs. 3, 4, and 5). As far as the authors are aware, our study is the first that attempts to use human perception and deep-learning techniques to evaluate playability at a city scale.

Given that the downtown area of many cities provides differential benefits to people based on age and gender, among other factors, more inclusive city planning, such as outlined by the 8 to 80 cities project,¹⁸ might be targeted at such areas. Liveliness and playability are not necessarily exclusive in downtown areas, and some urban planning approaches such as more parks and green space as well as reduced traffic volume and velocity could benefit both liveliness for adults and playability for children. While a negative correlation between playability and liveliness for downtown areas is inline with previous research, city-wide results may suffer from issues related to the underlying data. In particular, the question framed in MTurk during data collection for playability scores makes no distinction as to time of day when the scenes in the street view images would be used for play. In the case of a downtown area, for example, people scoring the images may assume that they should evaluate playability when downtown is busy, i.e. when it is most lively, when in fact playing and downtown activities such as dining frequently occur at different times. In such cases, lively and playable places may not actually be spatially exclusive, as long as the temporal difference is accounted for.

Similarly, lively scores may reflect vagueness in the question posed during data collection. In the place perception dataset, lively scores were obtained by asking participants to choose the most lively street view image from two presented street view images. Since no definition of lively was provided, different participants may have made different assumptions about liveliness, such as time of day considered and the demographics of the users (e.g. age, gender). Furthermore, a recent study by Yao et al. (2019) found that global-coverage urban perception datasets produced by machine learning techniques may not be very accurate at the local scale, casting some doubt on the veracity of local perception results in this study. For future research, we advise further exploration of how lively and playability scores vary geographically in various cities, and how they are tied to land use. The use of CBGs here as the spatial analysis unit may be obfuscating finer-grain land use, and as such a smaller spatial unit might be advisable for such research.

6.3. Broader implications

For playability, the city-scale, geographically-dependent analysis described here suggests that research on play can be done at larger scales than in traditional studies, and that the use of deep learning can allow researchers to utilize comprehensive underlying features from a much richer data source to describe neighborhood places. Generally, the use of street view images as a big data source for quantifying place perception at a city scale could be a cost-effective method for urban planners and geographers to quantify and describe places. Given the geographic variation in perception scores such as liveliness and playability, we suggest that such perception based metrics be used as a starting point for comprehensively describing places, at which point methods such as surveys and more in-depth evaluations can be used to describe places for the purposes of urban planning and development.

For sustainable city design, this study also provides valuable insights. According to the regression results, more green and beautiful space, greater proximity to open space, and lower speed limits may be associated with environments that are conducive to play. Such natural areas have many sustainability benefits to cities, and their positive impacts on playability provide yet more impetus for construction and maintenance of such areas. Similarly, reducing traffic speeds in cities could have the benefit of promoting sustainable forms of transportation such as biking and walking, while also benefiting playability.

¹⁵ <https://sf.gov.org/sfplanningarchive/zoning-maps>

¹⁶ <http://www.bostonplans.org/3d-data-maps/gis-maps/zoning-maps>

¹⁷ <http://www.seattle.gov/dpd/research/GIS/webplots/Smallzonemap.pdf>

¹⁸ <https://www.880cities.org/>

For the growth and development of children, constructing cities in such ways would have many benefits. Physically, children would have more opportunities to play and receive associated health benefits. Reducing traffic speeds would also provide them with safer places to play. Mentally, exposure to natural areas would afford children more focused and reduced anxiety (McCurdy et al., 2010).

With all of these benefits, governments and urban planners should consider playability when planning city development. These factors not only benefit growth and development in children, but also help create environmentally-friendly cities for people to live in.

6.4. Limitations and future work

Although we have proposed a computational framework for scoring neighborhood playability and analyzing its associations with geographical features, there is still room for improvement in the methodology used in this study. In particular, data bias may have impacted the results. Data bias is a common issue in the era of big data (Gao et al., 2017; Yang et al., 2017). Regarding bias from the labellers, 210 workers labelled all of the Boston images. No filters limiting which workers could accept the MTurk task were used. The percentages of male and female workers were 69% and 31%, respectively, while the average age of all workers was 34. MTurk workers have been shown to be more highly educated than U.S. adults in general (Hitlin, 2016). Given that more educated parents are less likely to allow their children to engage in unstructured free play (Aarts et al., 2012), the training data may be biased towards unplayable, which is consistent with the data distributions in Fig. 7. Overall, the task of labelling images for perception is inherently one that incorporates bias, as labellers are asked to assess the playability of places based on their own experiences. While education can potentially impact people's assessment of playability, perception is inherently based on subjective feelings, and thus will be dependent upon the labellers background, to some extent. For the questions to MTurk workers regarding playability, no distinction is made between playability for boys versus girls or children versus adolescents, both of which have been shown to impact the type and amount of play (Katzmarzyk et al., 2018). In order to tailor to the specific needs of different age groups, labelled datasets specific to certain age groups could be helpful in identifying playable and unplayable areas. Additionally, factoring in the perceptions of children may help make the model outputs better reflect the true playability of urban areas.

Regarding the data used in the regression analyses, we attempted to use the most up-to-date information available for all data layers. However, some layers may incorporate data from a number of years, depending on when the data was collected for a given location. In the case of the street view images, the range in date of collection is from 2007 to 2019, though the majority (more than 95%) are from 2018 and 2019. On such a time scale, the built environment is relatively stable, but to reduce fluctuations in values resulting from differences in collection date at various locations, we first average each data layer at the CBG level. We also note that street view images have directionality, such that in a given location street view images taken from different angles could receive different playability scores. Most notably, a street view image taken parallel to the road would likely score lower for playability since it has more pixels of the road, while a street view image from the same location but taken perpendicular to the road would capture more of the community landscape and thus might receive a higher score. Presumably, using the average playability score of the images from different angles at the same location mitigates this issue to a certain degree. In addition, future research might consider the standard deviations of playability at the same place when measuring the aggregate playability score.

Another future direction is to examine the relationships between land use patterns and playability. Previous studies have noted how relationships between the urban environment and play are specific to land use patterns, which are configured to serve different needs (Booth,

Pinkston, & Poston, 2005). Similarly, in our study, we found that the difference between playability and lively might also be influenced by urban functions. Hence, taking land use patterns into account may enhance our understanding of neighborhood playability.

7. Conclusion

Play has positive effects on childhood development and well-being, and is necessary for sustainable city development. In this study, we propose a theory-informed and computational framework that integrates big data and machine learning to evaluate place playability at a city scale. Specifically, we capture people's perception of place playability by employing Amazon Mechanical Turk to classify street view images. A deep learning model trained on the labelled data is then used to evaluate neighborhood playability for three U.S. cities: Boston, Seattle, and San Francisco. Finally, multivariate and geographically weighted regression models are used to explore how various urban features are associated with playability. We suggest that this method can be used as a cost-effective step for identifying areas of improvement for play at the city level, after which more studies and initiatives can be used to evaluate local needs. The high accuracy of spatial models in this study suggest that more research on play should use spatially explicit models, as variable importance for playability varies across space. In all cities, place perception variables are highly predictive of playability, suggesting that the place perception variables extracted from street view images can well capture various aspects of human perception in urban areas. Finally, in line with other studies, we find that higher traffic speeds and crime counts are negatively associated with playability, suggesting that both aspects be addressed by cities seeking to make their neighborhoods more playable. From the human perception perspective, higher scores for perception of beauty are positively associated with playability. Interestingly, a place that is perceived as lively may not be playable. Our research provides helpful insights for urban planning focused on sustainable city growth and development, as well as for research focused on creating nourishing environments for child development.

Author statements

Jacob Kruse: data collection and processing, analysis, visualization, result interpretation, writing.

Yuhao Kang: data collection and processing, analysis, result interpretation, writing.

Yu-Ning Liu: data collection and processing, visualization, writing.

Fan Zhang: research design, result interpretation, writing.

Song Gao: research design and conceptualization, analysis, result interpretation, writing.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Acknowledgements

The authors would like to thank the support from the Microsoft AI for Earth Program and the National Natural Science Foundation of China (No. 41971366, 41901321) to our study. Support for this research was partly provided by the University of Wisconsin–Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

References

- Aarts, M.-J., de Vries, S. I., Van Oers, H. A., & Schuit, A. J. (2012). Outdoor play among children in relation to neighborhood characteristics: A cross-sectional neighborhood observation study. *International Journal of Behavioral Nutrition and Physical Activity*, 9 (1), 98.

- Almanza, E., Jerrett, M., Dunton, G., Seto, E., & Pentz, M. A. (2012). A study of community design, greenness, and physical activity in children using satellite, GPS and accelerometer data. *Health & Place*, 18(1), 46–54.
- Bell, J. F., Wilson, J. S., & Liu, G. C. (2008). Neighbourhood greenness and 2-year changes in body mass index of children and youth. *American Journal of Preventive Medicine*, 35(6), 547–553.
- Booth, K. M., Pinkston, M. M., & Poston, W. S. C. (2005). Obesity and the built environment. *Journal of the American Dietetic Association*, 105(5), 110–117.
- Brussoni, M., Lin, Y., Han, C., Janssen, I., Schuurman, N., Boyes, R., ... Mäse, L. C. (2020). A qualitative investigation of unsupervised outdoor activities for 10-to 13-year-old children: "I like adventuring but I don't like adventuring without being careful". *Journal of Environmental Psychology*, 70, 101460.
- Dadvand, P., Villanueva, C. M., Font-Ribera, L., Martínez, D., Basagaña, X., Belmonte, J., ... Nieuwenhuijsen, M. J. (2014). Risks and benefits of green spaces for children: A cross-sectional study of associations with sedentary behavior, obesity, asthma, and allergy. *Environmental Health Perspectives*, 122(12), 1329–1335.
- De Nadai, M., Vieriu, R. L., Zen, G., Dragicevic, S., Naik, N., Caraviello, M., ... Lepri, B. (2016). Are safer looking neighborhoods more lively? A multimodal investigation into urban life. In *Proceedings of the 24th ACM international conference on multimedia* (pp. 1127–1135).
- Duarte, F., & Alvarez, R. (2021). *Urban play: Make-believe, technology, and space*. MIT Press.
- Dyment, J. E., & Bell, A. C. (2008). Grounds for movement: Green school grounds as sites for promoting physical activity. *Health Education Research*, 23(6), 952–962.
- Fjortoft, I., Löfman, O., & Halvorsen Thorén, K. (2010). Schoolyard physical activity in 14-year-old adolescents assessed by mobile gps and heart rate monitoring analysed by gis. *Scandinavian Journal of Public Health*, 38(5 suppl), 28–37.
- Fotheringham, A. S., Brunson, C., & Charlton, M. (2003). *Geographically weighted regression: The analysis of spatially varying relationships*. John Wiley & Sons.
- Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 107(6), 1247–1265.
- Gao, S., Janowicz, K., Montello, D. R., Hu, Y., Yang, J.-A., McKenzie, G., ... Yan, B. (2017). A data-synthesis-driven method for detecting and extracting vague cognitive regions. *International Journal of Geographical Information Science*, 31(6), 1245–1271.
- Geburu, 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(50), 13108–13113.
- Gilliland, J., Holmes, M., Irwin, J. D., & Tucker, P. (2006). Environmental equity is child's play: Mapping public provision of recreation opportunities in urban neighbourhoods. *Vulnerable Children and Youth Studies*, 1(3), 256–268.
- Ginsburg, K. R., et al. (2007). The importance of play in promoting healthy child development and maintaining strong parent-child bonds. *Pediatrics*, 119(1), 182–191.
- Grigsby-Toussaint, D. S., Chi, S.-H., & Fiese, B. H. (2011). Where they live, how they play: Neighborhood greenness and outdoor physical activity among preschoolers. *International Journal of Health Geographics*, 10(1), 66.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
- Helbich, M., Yao, Y., Liu, Y., Zhang, J., Liu, P., & Wang, R. (2019). Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in Beijing, China. *Environment International*, 126, 107–117.
- Henry, M. (1990). More than just play: The significance of mutually directed adult-child activity. *Early Child Development and Care*, 60(1), 35–51.
- Hitlin, P. (2016). *Turkers in this canvassing: Young, well-educated and frequent users* (p. 437). Pew Research Center.
- Hu, S., Xu, Y., Wu, L., Wu, X., Wang, R., Zhang, Z., Lu, R., & Mao, W. (2021). A framework to detect and understand thematic places of a city using geospatial data. *Cities*, 109, 103012.
- Huang, B., Zhou, Y., Li, Z., Song, Y., Cai, J., & Tu, W. (2020). Evaluating and characterizing urban vibrancy using spatial big data: Shanghai as a case study. *Environment and Planning B: Urban Analytics and City Science*, 47(9), 1543–1559.
- Hurvich, C. M., & Tsai, C.-L. (1993). A corrected akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis*, 14(3), 271–279.
- Jacobs, J. (2016). *The death and life of great American cities* (Vintage).
- Kalogirou, S. (2020). *lctools: Local correlation, spatial inequalities, geographically weighted regression and other tools. R package version 0.2–8*.
- 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(3), 261–275.
- Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., & Ratti, C. (2020). Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy*, 104919.
- Katzmarzyk, P. T., Denstel, K. D., Beals, K., Carlson, J., Crouter, S. E., McKenzie, T. L., ... Stanish, H., et al. (2018). Results from the United States 2018 report card on physical activity for children and youth. *Journal of Physical Activity and Health*, 15(s2), S422–S424.
- Kim, J.-H., Lee, C., & Sohn, W. (2016). Urban natural environments, obesity, and health-related quality of life among hispanic children living in inner-city neighborhoods. *International Journal of Environmental Research and Public Health*, 13(1), 121.
- Lambert, A., Vlaar, J., Herrington, S., & Brussoni, M. (2019). What is the relationship between the neighbourhood built environment and time spent in outdoor play? A systematic review. *International Journal of Environmental Research and Public Health*, 16(20), 3840.
- Lämmle, L., Ziegler, M., Seidel, I., Worth, A., & Bös, K. (2013). Four classes of physical fitness in german children and adolescents: Only differences in performance or at-risk groups? *International Journal of Public Health*, 58(2), 187–196.
- Laxer, R. E., & Janssen, I. (2013). The proportion of youths' physical inactivity attributable to neighbourhood built environment features. *International Journal of Health Geographics*, 12(1), 31.
- Lee, K. H., Heo, J., Jayaraman, R., & Dawson, S. (2019). Proximity to parks and natural areas as an environmental determinant to spatial disparities in obesity prevalence. *Applied Geography*, 112, 102074.
- Li, X., Zhang, C., & Li, W. (2015). Does the visibility of greenery increase perceived safety in urban areas? Evidence from the Place Pulse 1.0 dataset. *ISPRS International Journal of Geo-Information*, 4(3), 1166–1183.
- Loebach, J. E., & Gilliland, J. A. (2016). Free range kids? Using GPS-derived activity spaces to examine children's neighborhood activity and mobility. *Environment and Behavior*, 48(3), 421–453.
- McCurdy, L. E., Winterbottom, K. E., Mehta, S. S., & Roberts, J. R. (2010). Using nature and outdoor activity to improve children's health. *Current Problems in Pediatric and Adolescent Health Care*, 40(5), 102–117.
- Middel, A., Lukaszczk, J., Zakrzewski, S., Arnold, M., & Maciejewski, R. (2019). Urban form and composition of street canyons: A human-centric big data and deep learning approach. *Landscape and Urban Planning*, 183, 122–132.
- Montello, D. R., Goodchild, M. F., Gottsegen, J., & Fohl, P. (2003). Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation*, 3(2–3), 185–204.
- Nguyen, A., Borghese, M. M., & Janssen, I. (2018). Pedestrian traffic safety and outdoor active play among 10–13 year olds living in a mid-sized city. *Preventive Medicine Reports*, 10, 304–309.
- Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019). mgwr: A python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, 8(6), 269.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Quercia, D., O'Hare, N. K., & Cramer, H. (2014). Aesthetic capital: What makes London look beautiful, quiet, and happy?. In *Proceedings of the 17th ACM conference on computer supported cooperative work & social computing* (pp. 945–955).
- Richards, D. (2010). *Relationship between speed and risk of fatal injury: Pedestrians and car occupants*.
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In , vol. 57. *Proceedings of the 9th Python in science conference* (p. 61). Austin, TX.
- Spencer, C. J. (2006). *Children and their environments: Learning, using and designing spaces*. Cambridge University Press.
- Stevenson, M., Sleet, D., & Ferguson, R. (2015). Preventing child pedestrian injury: A guide for practitioners. *American Journal of Lifestyle Medicine*, 9(6), 442–450.
- Summerbell, C. D., Waters, E., Edmunds, L., Kelly, S. A., Brown, T., & Campbell, K. J. (2005). Interventions for preventing obesity in children. *Cochrane Database of Systematic Reviews*, 3.
- Tamis-LeMonda, C. S., Shannon, J. D., Cabrera, N. J., & Lamb, M. E. (2004). Fathers and mothers at play with their 2-and 3-year-olds: Contributions to language and cognitive development. *Child Development*, 75(6), 1806–1820.
- W.H.O. (2015). *Ten strategies for keeping children safe on the road*. Technical report. World Health Organization.
- Whitzman, C., & Mizrahi, D. (2012). Creating child-friendly high-rise environments: Beyond wastelands and glasshouses. *Urban Policy and Research*, 30(3), 233–249.
- Wridt, P. (2010). A qualitative gis approach to mapping urban neighborhoods with children to promote physical activity and child-friendly community planning. *Environment and Planning B: Planning & Design*, 37(1), 129–147.
- Yan, B., Janowicz, K., Mai, G., & Gao, S. (2017). From ITDL to Place2Vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In *Proceedings of the 25th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 1–10).
- Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big data and cloud computing: Innovation opportunities and challenges. *International Journal of Digital Earth*, 10(1), 13–53.
- Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, J., Wang, R., Wang, J., & Guan, Q. (2019). A human-machine adversarial scoring framework for urban perception assessment using street-view images. *International Journal of Geographical Information Science*, 33(12), 2363–2384.
- Yeh, A. G., Yue, Y., Zhou, X., & Gao, Q.-L. (2020). Big data, urban analytics and the planning of smart cities. In , vol. 0. *Handbook of planning support science* (pp. 179–198).
- Zhang, F., Fan, Z., Kang, Y., Hu, Y., & Ratti, C. (2021). "Perception bias": Deciphering a mismatch between urban crime and perception of safety. *Landscape and Urban Planning*, 207, 104003.
- Zhang, F., Zhang, D., Liu, Y., & Lin, H. (2018). 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. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning*, 180, 148–160.
- Zhou, H., He, S., Cai, Y., Wang, M., & Su, S. (2019). Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. *Sustainable Cities and Society*, 50, 101605.