Department of Land Economy

Environment, Law & Economics

UNIVERSITY OF CAMBRIDGE
Real Estate Research Centre

Working Paper Series

No. 2020-03

Title: Machine Learning, Architectural Styles and Property Values

Authors: Thies Lindenthal and Erik B Johnson

Affiliation: University of Cambridge and University of Alabama

Contact corresponding author: Thies Lindenthal, htl24@cam.ac.uk

Machine Learning, Architectural Styles and Property Values

Thies Lindenthal* (University of Cambridge) & Erik B. Johnson (University of Alabama)

12 May 2020

Abstract

This paper reveals the nuanced nature of demand side preferences for architectural style. We present evidence that architectural style is an important determinant of housing sales price and that these effects may be sensitive to neighborhood aesthetics and attenuated by new construction. We also introduce a general algorithm that scrapes housing photos and utilize a deep-learning based model to automatically classify homes by their architectural style, explicitly incorporating spatial dependencies in the exteriors of homes. Comparisons between classifications from the machine based model and expert humans illustrate ways to help detect and mitigate potential biases in image based machine learning methods.

*Corresponding author: Thies Lindenthal (htl24@cam.ac.uk). Replication files are available from the author's Github repository https://github.com/thies. Paul E. Glade and Lukas Heckmann-Umhau are thanked for exceptional research assistance. Carolin Schmidt, Peter Brummund, Mike Langen, Colin Lizieri, Franz Fuerst, Stanimira Milcheva, D'Maris Coffman, Norm Miller, Stephen Malpezzi, John Clapp, Pavel Krivenko and participants at the ASSA/AREUEA 2019 annual meeting and seminars at UCL/Bartlett and the Homer Hoyt Institute offered excellent feedback and critique on earlier versions of this paper.

Introduction

Understanding household preferences for architectural style is of increasing interest to both researchers and policy makers due to the perceived linkage between architectural style, home values, and housing supply. Britain's *Building Better, Building Beautiful Commission*, which advises the government on design choices for homes and neighborhoods, lists one of its primary aims as *"To make the planning system work in support of better design and style, not against it."* (The Economist 2018). Thus, matching the style of new housing to aggregate neighborhood preferences should ease the objections of incumbent neighborhood households and result in a more elastic housing supply. If the planning system considers neighborhood preferences for style, then increases in demand for housing in highly productive urban areas should result in relatively more building, less price pressure thereby promoting in migration and productivity growth.

The underlying policy assumption that links architectural style to housing supply elasticity is that households are more likely to support projects that match their taste preferences. An alternative explanation for incumbent homeowner's reticence to support new construction is that in the absence of new building, positive demand shocks are full capitalized into the housing prices of existing homes. If this is case, then objections based on style preferences may simply be a convenient way to increase the cost of new construction and minimize housing supply elasticity. In this paper, we rigorously test the assumption that households actually have preferences for architectural style. Our results indicate that 1) there is a sales price premia associated with a variety of architectural styles, 2) the price premia is sensitive to neighborhood aesthetics and 3) the style premia effect is found for existing structures and not for new buildings.

Building styles and property values

Buitelaar and Schilder (2017) find evidence on the link between architectural style and housing price Using architectural assessments by human experts, and show estimate a sizable premium of 5% for new buildings in the Netherlands that refer to traditional styles and a staggering 15% premium for new buildings that closely follow traditional shapes, facade composition and details. Tis study disentangles the architectural style from other unobserved characteristics such as building quality, differences in location, or year of construction. These controls are crucial, as earlier work has established that age and style variables tend to be highly correlated. Coulson and McMillen (2008) suggest a non-parametric estimator and establish a U-shaped age function and distinct price discounts for postwar and contemporary styles (vis-à-vis more historic styles). Francke and van de Minne (2017) investigate the depreciation of residential real estate in the Netherlands and decompose land versus structure values singling out the effect of "physical deterioration, functional obsolescence and and vintage effects". They find that buildings from the 1930s carry a strong price premium. Both of these results support the hypothesis that architectural style preferences could help determine housing supply elasticity.

Additional support for preferences over architectural styles comes from a large scale assessment of buildings' exteriors will allow for an analysis of the externalities of architecture. Buildings hardly ever stand in isolation and Ahlfeldt and Mastro (2012) investigate the influence a building's architecture exerts on its surroundings. They observe a positive price effect for residential buildings in the direct proximity of iconic homes by Frank Lloyd Wright in Oak Park, Illinois. A building's exterior does not need to be an architectural masterpiece to co-determine the value of other houses close by. Homogeneity of building shapes within street segments does influence property values. A similarly shaped neighboring building is value enhancing while proximity to a wildly different neighboring shape, everything else remaining equal, is detrimental to property values (Lindenthal 2017a).

Unfortunately, the traditional approaches chosen by e.g. Buitelaar and Schilder (2017) or Ahlfeldt and Mastro (2012) do not scale well. Each observation has to be classified into architectural styles by a human expert, which is time-consuming and costly, imposing an upper limit on the number of observations and the level of detail captured for each observation in any given sample. Combining Google Street View data with deep learning based classifiers offers a solution to this nexus as it captures the images of almost all buildings in many cities around the world at high level of accuracy and detail. By classifying a large sample of structures we are able to detect not only the architectural style of the building but also to characterize the neighborhood style. This allows for the identification of heterogeneity in taste preferences by neighborhood style class. The challenge we address in this paper is to extract and utilize building level information from this ubiquitous sensor, using freely available deep learning techniques.

Automated classification of architectural style

As Helbich et al. (2013) states, the benefit of using new methods to observe the built environment is that "essential determinants influencing real estate prices [which] are constantly missing and are not accessible in official and mass appraiser databases". This is certainly the case in our context, as understanding the impact of architectural style on housing prices is made difficult by the paucity of sales or assessment data that includes architectural style as a characteristic.

Recent work illustrates the potential of using street level imagery and machine learning classification for measuring previously unobserved characteristics of the urban environment. Naik et al. (2017) describes how neighborhood demographics may impact the physical appearance of neighborhoods. Gebru et al. (2017) use classified vehicle make and model information to predict income, race, education, and voting patterns at the precinct level. Glaeser et al. (2018) predict income in New York City. Naik, Raskar, and Hidalgo (2016) create a neighborhood safety based Streetscore which is shown to be highly correlated with neighborhood population density and household income. De Nadai et al. (2016) find that greenery and street facing windows contribute to a positive appearance of safety while Liu et al. (2017) evaluate the quality and upkeep of the built environment along Beijing's streets.

In contrast the block, street, or street-section level classifications used in the majority of previous studies that use automate image classification, Glaeser, Kincaid, and Naik (2018) push the level of observation from the to the individual *building level*. Utilizing images of buildings' exteriors collected from Google Street View[1], and to a lesser degree interior images from Zillow, they find that looks matter, at least in Boston: A one standard deviation improvement of a building's exterior is associated with an additional USD 70,000 in home value. Intuitively, the link between good looks and value is bi-directional: The appearance of buildings that went through foreclosure deteriorated significantly (Glaeser, Kincaid, and Naik 2018).

We follow the framework used by Glaeser, Kincaid, and Naik (2018) and focus on individual buildings as our unit of observation. This focus unlocks one of the main benefits of using mass collected street level imagery in economic research: property characteristics previously deemed "unobservable" can be directly observed in an accurate, objective, automatic and therefore be implemented in a more efficient, accessible and cost effective way.

Residential transaction data

Residential real estate transactions are public data in the UK, collected and published by the Land Registry (Land Registry 2017). The records include the date of transaction, price paid, street address, a classification of the property type (flat, detached, semi-detached, or terraced house), the estate type (freehold or leasehold) and an indicator for newly built properties. We select transactions from Cambridge which were recorded between January 1995 and October 2018, excluding any leaseholds, flats and properties classified as type "other", and sales with prices below £50,000 or in excess of £2,000,000. Linking recently taken images to sales data from 24 years assumes that houses have not changed their architectural styles – which seems reasonable. We are not aware of any conversions of just the exterior of buildings. In case of full redevelopments, we exclude any sales for a given address pre-dating the redevelopment, ensuring that a currently taken picture represents the house at all included transactions. Table 1 presents summary statistics for the sample.

– Insert Table 1 about here –

The Ordnance Survey *AddressBase* (Ordnance Survey 2017a) links street addresses to building outlines on Ordnance Survey maps (Ordnance Survey 2017b), which allows us to calculate the buildings floor plate (in

[1] https://www.google.co.uk/maps

m^2) and to estimate the building's volume from digital elevation models (Environment Agency 2015), as suggested by Lindenthal (2017b). We control for the location of each building by calculating the distance to the city center proxied by Great St. Mary's Church, and non-parametrically by using 69 indicator variables for each of the smallest census tracts (*Lower Super Output Areas*, LSOA) subdividing Cambridge (Office for National Statistics 2019). LSOAs typically have 1,000–3,000 residents and 400–1,200 households of comparable economic and socio-demographic characteristics (Office for National Statistics 2017).

Image Classification Methodology

In spirit of Helbich et al. (2013), we automate both the image collection and style classification processes which allows for the introduction of a novel, high quality and low cost hedonic characteristics to our model. Specifically, our algorithm collects images of individual UK buildings from Google Street View, classifies the depicted buildings using transfer learning on deep convolutional neural networks, combines the derived information with sales price data and, ultimately, estimates marginal prices for the estimated building characteristics. Additionally, we illustrate how machine classifications may be sensitive to photo quality by comparing machine classifications to those made by human experts.

Image collection

The first challenge when trying to collect building images is fundamental: How can one identify the building of interest correctly?

For the UK, and many other countries, the Google Street View API returns the coordinates of the nearest camera snapshot for a given location but fails to provide an accurate orientation and zoom-level of the camera needed to capture the front of the building. Fig. 1 presents a typical result from an address level search, showing a broad ensemble of buildings instead of singling out the building of interest (in this case, the partly captured terraced house at the very left margin)[2]. Thus, accurate image collection is a multistage process that utilizes a variety of independently collected but spatially related data.

– Insert Figure 1 about here –

First, we use Metadata queries to identify the finite set of locations where Google Street View has taken a snapshot across the study area. These queries return the nearest locations and dates of unique three dimensional panorama photos collected by a Google Street View car. Next, we assign each property to the nearest Street View photo location. Finally, we "aim" the Street View camera via panning and zooming at the front door of spatially defined parcels of interest. Each resulting image should be best 2D representation of the property of interest available in the 3D panorama[3].

[2]For major US cities, the accuracy of building image search results is higher and the Street View camera will pan towards the center of parcel the property sits on rather than pointing directly down the street. This makes the photo collection of studies such as Glaeser, Kincaid, and Naik (2018) much more straight forward.

[3]We test the robustness of our hedonic models to a variety of image 'quality' concerns in the results section of the paper

– Insert Figure 2 about here –

We illustrate the aiming of the camera from the nearest Google Street View 3D panorama point (green dot) based on (red dot) coordinates of a given building obtained from Ordnance Survey maps in Figure 2. Camera aiming is based on a viewshed algorithm that identifies which exterior walls are visible from the panorama point, ignoring any wall segments where the direct line of sight from the panorama point is obstructed by other buildings. From this analysis, we are able to estimate the camera bearing (green line) and zoom factor, based on the fan of the lines of sight (in blue).[4] We apply this algorithm to automatically build a set of building frontage images for approximately 48,000 properties in Cambridge (UK)[5]. We download all pictures at the highest resolution offered by Google's Street View API[6] (640x640 color pixels).

Image classification

The next step is to build a model that classifies the architectural style observed in each of the collected Street View images. We do this through the use of 1) an expert-classified training data set and 2) applying transfer learning to an existing image classification model.

Training Data

The first step in building the model to assign house images to architectural styles is to construct a training data set. The training data set should contain a variety of examples of each different style. These examples can be then be used to teach the model to recognize architectural style in a variety of settings, angles and in the presence of visual confounders such as cars, trees, building colors and non architectural walls. For the study area, final year students from the Architecture Department at the University of Cambridge identified seven distinct architectural style classifications:

- *Georgian* (c1714–1837) houses feature key characteristics such as sash windows, fan lights above doors, the use of stucco on facades, often wrought work grilles, railings etc.
- In the *Early Victorian* era (c1837–c1870s), a growing taste for individualized embellishment led to the development of elaborate features such as carved barge boards or finials. The development of sheet glass led to sash windows becoming more affordable, and, increasingly, wider.

[4]Obstructed lines of sight due to greenery, fences, garden walls or large vehicles cannot be detected from the Ordnance Survey maps. We therefore use a first stage image classification procedure to identify if the building image is obstructed and we proceed the second-closest location if the line of sight is obstructed. Code necessary to replicate the full image capture including metadata capture and viewshed based camera aiming is available from the authors' GitHub.com code repositories.

[5]The Ordnance Survey maps do not provide any use type classifications so we collect images on all buildings that have a ground plate between 40 and 250 m^2. This method, however, could easily be adapted to photograph buildings in locations for which high-resolution building level maps are not available by using LIDAR based building outlines instead, which would be in the spirit of Glaeser et al. (2018).

[6]https://developers.google.com/maps/documentation/streetview/intro

- In the *Late Victorian* era (c1870s–1901), bay windows became more and more widespread, and increasingly substantial. Stylistic movements such as the Queen Anne revival style contributed richly ornamental details to the formal repertoire employed by designers. Stained glass became more popular.
- *Edwardian* architecture (1901-1910) tends to be less ornate than late Victorian architecture.
- The *Interwar* period (1918–1939) saw the cost of building construction fall, amidst a drive to provide better housing for the working classes. New housing types were being favored.
- The *Postwar* (1950-1980) styles continued on this path, with an embrace of high-rise as well as low rise housing. Facades vary greatly between brick, tiling, pebbledash and render.
- The cut-off year for our *Contemporary* era to begin is 1980. *Revival* style-homes are contemporary buildings trying to emulate historic, mostly Victorian, architecture.

To build the training data set, the experts then assigned a sample of 25,000[7] images to their corresponding architectural style[8].

Next we create a stratified training samples of 600 buildings from each category[9]. This leaves a sufficiently sized out-of-sample verification data set of 21,000 homes.

Training the model

We leverage recent advances in transfer learning to enable the use of smaller training datasets and much lower computational burden that traditional deep learning models. Transfer learning works by freezing the parameters in a pre-trained image recognition model, removing some of the final steps in the model, and then training a new model on the vector of outputs to produce new classifications[10]. For our purposes, we use the *Inception-v3* deep convolutional neural network (Szegedy et al. 2015). *Inception-v3* has been trained for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[11], which evaluates image classification and object detection algorithms for a wide range of objects. The pretrained classifications would allow us to identify pets, vehicles or people on the pictures – assessing architectural style, however, is beyond the canned classifiers' capabilities. Thus we strip off the last steps of the model which then outputs a 2048-dimensional feature vector for each picture. This feature vector basically describes the outline shapes and locations, colors, that were important for the ImageNet classficiations. Esentially, the

[7]The full set of classified images are available for download at the authors' websites.

[8]This is a much larger sample than is actually needed to build our training data set. Due to the use of transfer learning, each category requires less than 250 observations to reach saturated training accuracy levels. We greatly exceed this number for the purpose of this paper so that we can compare the out-of-sample convolutional neural network predictions to the architects' classifications. This allows us to examine the power and size of the assignment tests. In addition having both human and machine classification for a large sample of the data allows for robustness checks on the machine comparisons.

[9]This is true for all but the *Georgian* category, which is the least common style in Cambridge and for which we can only sample 330 examples.

[10]https://www.tensorflow.org/tutorials/images/transfer_learning

[11]http://image-net.org/challenges/LSVRC/

Inception-V3 model works as a sophisticated dimension reduction engine that we can then use to train a new, more simple model to detect architectural style[12].

We next train our model using a simple multinomial classifier comprising of:

- an input layer the size of the feature vectors, e.g. 2048 or 4096, respectively,
- one dense layer (relu) half the size of the input layer, one subsequent dropout (rate 0.5) layer,
- one dense layer (relu) a quarter the size of the input layer, one subsequent dropout (rate 0.5) layer,
- and the final dense output layer with softmax activation.

All classifiers are implemented using the Keras/Tensorflow APIs[13]. The computational burden of this rather shallow model design is modest.

We implement an ensemble learning approach: For 100 randomly drawn data sets, all stratified on style, we train two models each: The first is based on the building specific feature vectors (2048 dimensions) while the second also incorporates the feature vector of the nearest neighbor (combined 4096 dimensions). The architectural styles of neighboring buildings tend to be similar, which means that incorporating information from the nearest neighbors can improve the classification, especially in cases where trees, fences or large vehicles obstruct the view or where the front of a building cannot be classified with high certainty. The image of the nearest neighbor provides a "second opinion" that is spatially correlated with the observation of interest. Finally, we classify all buildings from the Ordnance Survey but not used in training or evaluating the model (truly out-of-sample).

When classifying a building picture, the softmax activation layer returns a vector of style-scores, each between 0 to 1, that jointly sum up to 1. We select the style with the highest score as the best estimate.[14]

There are two variables are obtained from the distribution of classifications per observation: The ensemble classifier arrives at predictions by majority vote. In addition, a confidence measure is computed as the Herfindahl index of classifications by ensemble member models. If, for instance, 20 models classified a building as *Contemporary* but 80 others as *Revival*, the ensemble majority vote is *Revival*. The corresponding Herfindahl score is calculated as $(20/100)^2 + (80/100)^2 = 0.68$.

[12]Glaeser, Kincaid, and Naik (2018) rely on an different ILSVRC competitor, *Resnet-101* (He et al. 2016). They reduce the extracted feature vectors to lower dimensionality (1024 to 100 dimensions) based on principal component analysis (PCA). We follow a different strategy and double the dimensionality (4096) by including the feature vector of the closest building in the sample, which, in most cases, is the direct neighbor. The feature vectors of neighbors have been collected in exactly the same way as those for other buildings in the sample. Doubling up allows us to model spatial dependencies in building styles, similar to spatially correlated land cover classifications in Ghimire, Rogan, and Miller (2010).

[13]Keras: https://keras.io/, Tensorflow: https://github.com/tensorflow/tensorflow

[14]By not excluding observations where multiple scores are vying for the top rank we retain as many observations as possible but risk a higher misclassification rate. We examine the role of model 'confidence' in the Machine Classification Diagnostics section.

Hedonic regression setup

Merging the machine's classification with sales data (Land Registry 2017) we estimate a hedonic regression equation that establishes marginal prices for the building style (similar to Moorhouse and Smith 1994; Asabere, Hachey, and Grubaugh 1989; Vandell and Lane 1989; Fuerst, McAllister, and Murray 2011; Plaut and Uzulena 2006), among other characteristics:

$$ln(Price_i) = \alpha + \beta\mathbf{X_i} + \delta\mathbf{Style}_i + \eta\mathbf{StyleNeigh}_i + \iota\mathbf{Style}\cdot\mathbf{StyleNeigh}_i + \gamma\mathbf{Y}_i + \lambda\mathbf{Loc}_i + \epsilon_i \quad (1)$$

Here, the natural logarithm of sales prices is explained by a linear combination of hedonic attributes described in vector $\mathbf{X}$, vectors of year $\mathbf{Y}$ and neighborhood $\boldsymbol{Loc}$ dummy variables and the building's estimated **Style** and the prevailing styles of other buildings in the direct proximity (**StyleNeigh**). $\mathbf{Style}\cdot\mathbf{StyleNeigh}$ is a vector of interaction terms for the building's and the neighborhood's dominant style. The intercept is denoted by α while $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, $\boldsymbol{\eta}$, $\boldsymbol{\iota}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are vectors of regression coefficients. ϵ is the IID error term. Heteroscedasticity robust standard errors will be reported.

Are buildings with different appearances imperfect substitutes catering to multiple groups of households with distinct style or vintage preferences? Buitelaar and Schilder (2017) indicate that any premium for an architectural style must stem from either differences in construction prices (which they do not find in their Dutch sample) or from supply constraints, as new construction potentially does not capture the demand for traditional styles. For Cambridge, new supply will inevitably be of either contemporary or revival style as historic vintages (and their styles), by definition, are not supplied any more. Estimating Eq. 1 for a subset of newly constructed buildings will show whether construction prices or supply constraints for new homes built according to different architectural styles prevail – if too few vernacular buildings were built, prices should reflect such a shortage. Singling out new buildings crucially controls for the otherwise unobserved age of these buildings, which is tightly intertwined with their aesthetics.

Results

Machine Classification Diagnostics

To measure how well our machine based classifiers compare to the human-experts, we tabulate confusion matrices and perform a regression. in Table 2. Panel A illustrates the raw counts, percents and accuracy statistics for each architect defined class. For example, there are 284 homes which are classified as Georgian by both the Architects and the Machine and 79% of all expert classified Georgian are correctly identified by the model. 50 homes are classified as Georgian by the experts but misclassified at Early Victorian by the machine. The machine classifier performs best on Early Victorian, Late Victorian/Edwardian and

Interwar and performs the worst at detecting Georgian, Contemporary and Revival styles. The reason for the this variation is two fold. First, there are relatively few Georgians in the training data set (and in the overall population) so variations it is harder to control for visual confounders and the precision denominator over reports false positives relative to other classes due to its relatively lower sample size. Secondly, both Revival and Contemporary housing borrow architectural details from the styles common in previous eras.

In Table 3 we regress the style classifications of the human-expert architects, captured by binomial variables $D_{TrueVint}$, on a vector of hedonic variables $\mathbf{X}$, vectors of year $\mathbf{Y}$ and neighborhood $\boldsymbol{Loc}$ dummy variables and the building's **Style**, estimated both with and without spatial dependencies. The intercept is denoted by α while $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are vectors of regression coefficients, and ϵ is the error term.

$$logit(D_{TrueVint,i}) = \alpha + \beta\mathbf{X_i} + \delta\mathbf{Style}_i + \gamma\mathbf{Y}_i + \lambda\mathbf{Loc}_i + \epsilon_i \tag{2}$$

The logit regressions are estimated by generalized least squares. For a well-performing classifier, the δ coefficients will be statistically significant and, more importantly, differences in the Akaike Information Criterion (AIC) allow for comparisons across the suggested ML classifiers.

Will the resulting classifiers be biased with respect to property values, picking up correlated but irrelevant cues such as greenery, upkeep, or cars brands instead of building exteriors? Are, for instance, more valuable houses more likely to be classified as the popular *Late Victorian* style while starter homes are disproportionally more likely to be classified as e.g. *Early Victorian*? We expect that in cases in which the ensemble provided a high confidence score, building features are clearly detectable on the pictures. For close calls, however, where models within the ensemble cannot agree, other factors might tilt the scale towards an incorrect class. A sizable difference in estimated marginal prices for machine classifications versus the estimates for architects' classifications in the subsequent hedonic regression (Eq. 1, below) would indicate such a bias.

Almost 16,000 of the 25,000 buildings for which we have an architect's classification could be matched to the 23,768 sales transactions that have been recorded for Cambridge between 1995 and 2018. The remaining 9,000 houses have not been sold since 1995 - but the information on their architectural style is still valuable for our measures of neighborhood styles. Thus, we observe both the machine's and the architects' classification for the majority of buildings in the data set.

– Insert Table 2 about here –

The agreement between classifications by the machine and by the architects is reassuringly strong. When relying on feature vectors of the buildings and their closest neighbors, 67 to 84 percent of true classifications are matched with correctly estimated labels (Table 2, Panel A). The recall rates are especially high for older styles: For *Georgians*, it is 79 percent, *Early Victorian* 81 percent, *Late Victorian/Edwardian* 78 percent, *Interwar* 84 percent, respectively. For more recent *Postwar* buildings, the rate drops to 71 percent, for *Contemporary* to 72 percent and for *Revival* to 67 percent.

A visual inspection of classifications is additionally reassuring. Figure 3 displays the pictures which carry the highest out-of-sample scores for each style. The model is clearly able to differentiate based on small cues, even when only parts of the facade are captured on the picture.

The Herfindahl scores for classifications vary substantially across styles, as classified by the architects (Table 3), indicating that the models within the ensemble, for instance, agree more often on a building being *Georgian* than finding consensus for other styles. For *Contemporary* or *Revival* styles, individual classifications differ most. All off-diagonal elements in the lower panel of Table 3 are negative, suggesting that, in general, the Herfindahl index is a good predictor of misclassified images.

A closer inspection of the misclassified images sheds some light on the limits of automatically collected street level imagery. What do misclassified images have in common? Off-the-shelf object detection algorithms can identify broadly defined objects such as trees, vehicles, houses, doors, or windows without any additional training. Using an Inception/Resnet object detection model trained on Open Images[15], we calculate the share of the image area taken up by cars or trees as a measure of view obstruction. The more image area is showing buildings and, importantly, windows the more meaningful input an automatic classifier can draw from, resulting in better classifications. In addition, images not taken at an optimal angle or zoom factor are detected by calculating the distance between the bounding box for the largest building detected on the image and the center of the image: For well-centered images, this offset will be small.

Table 4 presents the mean values for the image quality variables. Table 5, Panel A, compares the mean values for views obstructed by trees or vehicles for correctly (on diagonal) and incorrectly classified images. In most cases, the off-diagonal values are positive, showing that more obstacles in the view of sight correlate with a higher likelihood of misclassifications. Similarly, if more of the image area shows a house (Panel B), misclassifications are reduced (negative differences). This effect is stronger for windows, which appear to give valuable cues about a house's style (Panel C). Finally, houses that are not in the center of the building tend to be misclassified more frequently (positive differences in Panel D). In sum, bad quality images lead to misclassifications.

– Insert Table 3, Table 4 and Table 5 about here –

The number of misclassifications decreases decidedly when considering observations with high ensemble

[15] In this case, "faster rcnn inception resnet v2 atrous oidv2", available from https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md

confidence only: The majority vote based on the most confident two-thirds of machine classifications only are more in line with the architects' classifications, as higher recall, precision and F_1-scores in Table 2, Panel B, show. The improvement is most evident in the precision rates. For the relatively rare categories of *Georgian* and *Revival* about the same number of buildings classified as *Georgian/Revival* in Panel A actually belong to a different style, according to the human experts. The precision shoots up from 0.50 to 0.70 for *Georgian* and from 0.52 to 0.82 for *Revival* when excluding the not so confident lower third of predictions in Panel B.

The majority of misclassifications is concentrated in temporally adjacent eras: 10 percent of *Late Victorian/Edwardian* buildings are labelled as *Early Victorian* or 18 percent of *Postwar* buildings are erroneously regarded as stemming from the *Interwar* period. Hardly ever is a contemporary building mistaken for a historic home.

– Insert Figure 3 about here –

Overall, the spatial softmax classification incorporating the nearest neighbor's feature vector outperforms the base classifier: pairwise differences of the F_1-scores from the spatial and the base classifier are clearly positive (Fig. 4). AIC values for 28 binomial models based on Eq. 2 decrease as hedonic characteristics, location and time dummies are augmented by the base and spatial ML predictor (Table 6). For all building styles, the AIC are lowest for the spatial ML predictor. In sum, the spatial classifier performs better for UK style data than a classifier based on single buildings only.

– Insert Table 6 and Figure 4 about here –

Hedonic regression estimates

Successively, the estimated coefficients from 8 different versions of the hedonic regression specified in Eq. (1) are reported in Tables 7 and 9. For all models, the hedonic control variables show the expected signs: Negative coefficients for the relative distance to the city center, discounts for terraced homes and semi-detached homes relative to detached houses, positive elasticities for building floor plate and building volumes and a price premium for new buildings compared to second-hand homes. Year and neighborhood dummies control for time effects and local amenities but their coefficients are not reported due to space constraints. The combination of location dummies and the distance to the city center measure controls for proximity to the city center *within* each neighborhood.

– Insert Table 7 about here –

The first column in Table 7 presents the estimated regression coefficients for the architects' classifications. The base style *Contemporary* is more expensive than almost all other styles, which show negative coefficients that are significantly different from 0. A clear pecking order appears: *Georgian* and *Revival* buildings demand the highest premium (+0.04), followed by *Late Victorian/Edwardian* (+0.01), *Contemporary* (base), *Early Victorian* (−0.13), *Interwar* (−0.15) and *Postwar* (−0.22).

The sample of images classified by architects is not necessarily representative of the overall population so we re-estimate the model using machine classifications for *all* buildings for which we were able to extract usable pictures from Google Street View. Overall, the style coefficients become more positive relative to the base, implying that our automatic classification is biased and tends to classify bland or low quality buildings to the *Contemporary* style. Reversely, more attractive or higher quality buildings are disproportionately often misclassified as *Late Vic./Edw.*, pushing up the corresponding coefficient to 0.11. This bias vanishes in the third model, which uses only confidently estimated machine styles (Column 3).

Models 4–6 follow the same logic as 1–3 but are estimated for *new* buildings only, which are either in the contemporary or the revival style. Basically this contrasts houses which might *look* different from the street, having either contemporary or revival facades, but which are all modern homes at their core. Differences in e.g. materials, floor plans, green space and gardens are minimal. For newly built homes, no effect of *Revival* architecture on price can be found for high confidence ML estimates, neither for classifications by machines nor by architects. After controlling for location, building characteristics and quality, buyers show no willingness to pay a premium for *Revival* architecture (Columns 4 and 6).

Introducing a variable capturing the most prominent style of buildings on the same street and within 100 m (**StyleNeigh**) does not change the pecking order of building styles much (Table 7, Column 7). Neighboring styles are compared to the base case of a *Contemporary* house in a *Contemporary* street. The negative coefficients on dissimilar neighboring styles support a finding by Lindenthal (2017b) that shows that a harmonious match of a building's shape with its direct environment leads to a price premium – or a discount in case of shape mismatches. In relative terms, *Postwar* buildings are the least popular neighbors. The difference between *Revival* and *Contemporary* neighbors is not statistically significant. We fail to find evidence in favor of positive externalities from revival architecture. Everything else equal, buildings within ensembles of *Contemporary* generally do not achieve higher transaction prices than buildings in more historic areas (as Table 10 will confirm later).

Table 9 presents the ι interaction coefficients (Eq. 1) for style-neighboring style combinations with sufficient numbers of observations (see Table 8). The last row suggests that a general claim that *Revival* neighbors exert positive externalities and increase property values relative to *Contemporary* surroundings cannot be confirmed empirically. Again, we find no statistically significant difference.

– Insert Table 8, Table 9 and Table 10 about here –

In Table 10, the combined effect of building styles on property values is calculated by adding up the direct, neighborhood and interaction effects (Tables 7 and 9). When filling in lots in historical neighborhoods, however, a premium for *Revival* facades over modern designs can be observed (differences in last two columns of Table 10). For large scale new construction, however, *Revival* buildings surrounded by other historicizing buildings do not sell for more than *Contemporary* buildings in a more modern setting.

Conclusion

Does the the widely-held belief of home-buyers valuing traditional building styles more than contemporary styles hold up empirically? After accounting for building quality and location, which tend to be better for older vintages, no evidence for such a premium emerges from data on *newly built* homes. Observing the real-life choices made by home-buyers is more informative than any debate of aesthetics fueled by newspaper columnists, think tanks or ideological beliefs in general.

For newly constructed homes, we do not find a positive price effect of revival architecture on neighboring homes. Objecting to new construction might be in the interest of existing home-owners as it limits supply and thereby drives up the value of their homes. Objections based on architectural preferences are not supported by market data.

On the more technical side, this paper offers four contributions: First, it introduces an algorithm that collects pictures of *individual* buildings from Google Street View. Earlier work has not achieved this level of detail and was, at least in the UK, limited to street sections only. A large-scale application of automatic classification of individual buildings' characteristics using Google Street View has potential not only in the UK. The image collection and classification method can easily be ported to other study areas which have existing Street View data and either LIDAR based building outlines or high resolution satellite images. In a follow-up project, we are working on an improved classification workflow in which a customized object detection model is able to recognize specific building outlines without the need for additional maps or other data.

Second, we developed a new database of 25,000 building pictures that have been classified by architecture experts into relevant architectural styles. We subsequently trained a neural network classifier to automatically classify all residential buildings of a mid-sized English city into architectural styles. The suggested classifier is trained on feature vectors of buildings and their nearest neighbors to exploit spatial correlation in observed classifications.

The large ground truth data set allows for a comparison of human expert versus machine classifications. Poor image quality, for instance obstructed views or the lack of informative features such as windows, is correlated with misclassifications. Importantly, for cases in which the ML classifiers are relatively indecisive between classes, visual cues that are correlated with the exterior's style but not part of a building's architecture (for instance trees or cars) bias the estimates. For these less certain cases, more valuable homes are systematically more likely to be classified as e.g. *Late Victorian/Edw.* and less likely to be considered *Contemporary.* A follow up study could investigate which elements in the building pictures lead to the misclassifications (Ribeiro, Singh, and Guestrin 2016). Given the danger of systematic biases, one should remain wary of ML estimates derived from data sets sizable enough to train models – but too small to investigate any biases.

Third, we explore how spatial dependencies in image classifications can be exploited in deep neural networks. Including information from neighboring houses improves the predictive power of the image classifier significantly.

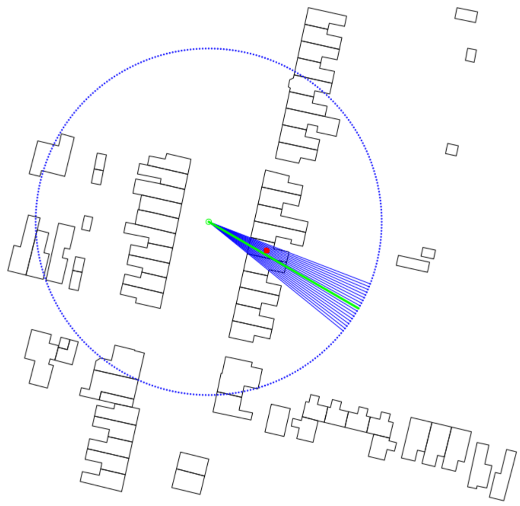
The last contribution is more practical. We investigate how prediction accuracy is influenced by image quality and find the obvious: Obstructions in the line of sight such as trees or cars make pictures of houses less informative. Future research could utilise building images taken from multiple angles to circumvent obstacles and to arrive at more reliable classifications. Street-level images are a excitingly rich data source for urban economics and real estate research – it just takes a bit of effort to tap into it.

Tables and Figures

Figure 1: ‘Street View’, not ‘Building View’ – Identification of buildings remains a challenge in the UK

Notes: For the UK, the Google Street View API returns the coordinates of the nearest camera snapshot for a given location but fails to provide an accurate orientation and zoom-level of the camera needed to capture the front of the building exactly. In this typical example, the building of interest is only partially shown at the very left margin of this result. *Image source*: Google Street View.

Figure 2: Image Collection on Google Street View: Camera Direction and Zoom



Notes: We first look up the nearest Google Street View panorama point (green dot) based on the centroid (red dot) coordinates of a given building obtained from Ordnance Survey maps. A viewshed analysis identifies which exterior walls are visible from the panorama point, ignoring any wall segments where the direct line of sight from the panorama point is obstructed by other buildings. The camera bearing (green line) and zoom factor are based on the angle of the most outer lines of sight (blue lines).

Table 1: Summary statistics residential property transactions

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price	23,768	253,646.20	170,987.20	10,000	127,000	326,374.8	1,000,000
Year	23,768	2,004.89	6.59	1,995	1,999	2,010	2,018
ln(volume)	23,768	286.00	188.14	0.00	215.50	370.89	1,926.45
ln(area)	23,768	63.25	28.44	4.94	45.74	72.46	1,058.70
ln(dist. city center)	23,768	2,477.63	1,007.55	112.78	1,652.39	3,166.05	5,023.62
New	23,768	0.04	0.20	0	0	0	1
Type: detached	23,768	0.13	0.34	0	0	0	1
Type: semi-detached	23,768	0.35	0.48	0	0	1	1
Type: terraced	23,768	0.52	0.50	0	0	1	1
Georgian	23,768	0.03	0.16	0	0	0	1
Early Vic.	23,768	0.15	0.36	0	0	0	1
Late Vic./Edw.	23,768	0.21	0.40	0	0	0	1
Interwar	23,768	0.32	0.47	0	0	1	1
Postwar	23,768	0.23	0.42	0	0	0	1
Contemporary	23,768	0.03	0.17	0	0	0	1
Revival	23,768	0.04	0.20	0	0	0	1
Neigh: Georgian	23,768	0.02	0.12	0	0	0	1
Neigh: Early Vic.	23,768	0.14	0.35	0	0	0	1
Neigh: Late V./Edw.	23,768	0.22	0.41	0	0	0	1
Neigh: Interwar	23,768	0.32	0.47	0	0	1	1
Neigh: Postwar	23,768	0.25	0.43	0	0	1	1
Neigh: Contemporary	23,768	0.02	0.15	0	0	0	1
Neigh: Revival	23,768	0.03	0.18	0	0	0	1

Notes: Summary statistics for sample of 23,768 residential real estate transactions for the city of Cambridge (UK) between 1995 and 2018 where buildings could be matched with Google Street View images. The buildings floor plate (in m^2) is based on OS maps and the buildings' volumes are estimated from digital elevation models (Lindenthal, 2017b). We control for the location of each building by calculating the distance to the city center proxied by Great St. Mary's Church, and non-parametrically by using 69 location dummies (based on LSOA). Building styles based on machine classifications.

Table 2: Confusion matrix for classification based on images of property and its nearest neighbor

Panel A: Model based on individual building and nearest neighbor

Machine	*Architects*						
	Georgian	Early Vic.	Late V./Edw.	Interwar	Postwar	Cont.	Revival
Georgian	284	109	77	38	14	23	22
Early Vic.	50	1755	427	86	86	56	34
Late V./Edw.	10	172	3260	213	56	21	29
Interwar	10	46	254	5884	997	56	54
Postwar	1	17	48	514	3914	74	40
Cont.	3	50	63	101	333	855	69
Revival	3	29	45	139	145	98	501
Georgian	79%	5%	2%	1%	0%	2%	3%
Early Vic.	14%	81%	10%	1%	2%	5%	5%
Late V./Edw.	3%	8%	78%	3%	1%	2%	4%
Interwar	3%	2%	6%	84%	18%	5%	7%
Postwar	0%	1%	1%	7%	71%	6%	5%
Cont.	1%	2%	2%	1%	6%	72%	9%
Revival	1%	1%	1%	2%	3%	8%	67%
Recall	0.79	0.81	0.78	0.84	0.71	0.72	0.67
Precision	0.50	0.70	0.87	0.81	0.85	0.58	0.52
F_1-score	0.61	0.75	0.82	0.82	0.77	0.64	0.59

Panel B: Model based on individual building and nearest neighbor, high confidence only

Georgian	122	31	9	4	2	1	5
Early Vic.	30	1,926	324	19	40	15	7
Late V./Edw.	0	105	2,941	68	19	0	9
Interwar	6	6	81	3,230	369	10	14
Postwar	0	3	5	96	1,831	9	6
Cont.	0	8	7	4	33	236	18
Revival	3	2	0	14	10	4	154
Georgian	76%	2%	0%	0%	0%	0%	2%
Early Vic.	19%	93%	10%	1%	2%	6%	3%
Late V./Edw.	0%	5%	87%	2%	1%	0%	4%
Interwar	4%	0%	2%	94%	16%	4%	7%
Postwar	0%	0%	0%	3%	80%	3%	3%
Cont.	0%	0%	0%	0%	1%	86%	9%
Revival	2%	0%	0%	0%	0%	2%	72%
Recall	0.76	0.93	0.87	0.94	0.79	0.86	0.72
Precision	0.70	0.82	0.94	0.87	0.94	0.77	0.82
F_1-score	0.73	0.87	0.90	0.90	0.86	0.81	0.77

Panel C: Model based on individual buildings only

Georgian	267	121	110	46	33	36	17
Early.Vic.	50	1673	438	112	107	63	52
Late V./Edw.	19	183	3099	272	86	33	44
Interwar	12	71	335	5786	1419	84	82
Postwar	2	24	56	462	3364	75	38
Cont.	5	63	82	141	366	778	79
Revival	6	43	54	156	170	114	437
Georgian	74%	6%	3%	1%	1%	3%	2%
Early.Vic.	14%	77%	10%	2%	2%	5%	7%
Late V./Edw.	5%	8%	74%	4%	2%	3%	6%
Interwar	3%	3%	8%	83%	26%	7%	11%
Postwar	1%	1%	1%	7%	61%	6%	5%
Cont.	1%	3%	2%	2%	7%	66%	11%
Revival	2%	2%	1%	2%	3%	10%	58%
Recall	0.74	0.77	0.74	0.83	0.61	0.66	0.58
Precision	0.42	0.67	0.83	0.74	0.84	0.51	0.45
F_1-score	0.54	0.72	0.78	0.78	0.70	0.58	0.51

Notes: Cross-tabulation of out-of-sample predictions by the machine versus the architects' classification. *Recall* is the share of buildings from an architects' category being predicted correctly (diagonal in mid panel) and *Precision* is the share of buildings predicted to belong to a category that are indeed from that category. F_1-scores are the harmonious mean of Precision and Recall: F_1*-score = 2 Recall * Precision / (Recall + Precision).*

Table 3: Prediction certainty: Herfindahl index from ensemble model

	Architects						
	Georgian	Early Vic.	Late V. /Edw.	Interwar	Postwar	Cont.	Revival
Mean Herf.	0.88	0.80	0.79	0.77	0.68	0.70	0.70
Machine	*Difference from correct classifications (diagonal), t-stats in parenthesis*						
Georgian	–	-0.21 (-9.87)	-0.28 (-13.11)	-0.33 (-9.95)	-0.27 (-6.30)	-0.25 (-7.97)	-0.25 (-7.54)
Early Vic.	-0.22 (-6.28)	–	-0.19 (-19.20)	-0.31 (-13.70)	-0.21 (-10.05)	-0.21 (-7.53)	-0.15 (-3.84)
Late V./Edw.	-0.36 (-5.89)	-0.24 (-18.03)	–	-0.26 (-21.88)	-0.23 (-9.08)	-0.24 (-4.96)	-0.16 (-4.33)
Interwar	-0.12 (-1.55)	-0.32 (-12.77)	-0.24 (-19.87)	–	-0.13 (-22.67)	-0.25 (-12.58)	-0.16 (-6.49)
Postwar	-0.51 (-33.71)	-0.30 (-6.24)	-0.36 (-13.13)	-0.26 (-34.82)	–	-0.22 (-10.34)	-0.24 (-7.70)
Cont.	-0.44 (-3.31)	-0.37 (-17.78)	-0.38 (-17.63)	-0.39 (-25.39)	-0.18 (-15.67)	–	-0.19 (-7.71)
Revival	-0.09 (-0.42)	-0.40 (-20.64)	-0.41 (-22.54)	-0.31 (-19.71)	-0.28 (-24.06)	-0.22 (-12.40)	–

Notes: For each ensemble classification i, we derive the Herfindahl scores as the sum of the squared share of votes from individual models for each of the 7 styles s received, calculated as: $Herf_i = \sum_{s=1}^{s=7}(votes_s/votes_{all})^2$. A high score indicates high levels of consensus within the ensemble. The Herf. scores tend to be lower at off-diagonal cells, indicating lower consensus for misclassifications.

Table 4: Mean Values for image quality variables, by style

Architects' classifications	Share Blocked	House Area	Window Area	Image Offset
Georgian	0.07	0.81	0.15	0.07
Early Vic.	0.06	0.86	0.16	0.05
Late V./Edw.	0.11	0.84	0.17	0.08
Interwar	0.19	0.70	0.11	0.12
Postwar	0.16	0.64	0.07	0.11
Cont.	0.09	0.72	0.11	0.09
Revival	0.10	0.78	0.11	0.09

Notes: Using an Inception/Resnet object detection model trained on Open Images, basic object on the images are detected and the share of the image area taken up by cars, trees, buildings and windows are calculated. Images not taken at an optimal angle or zoom factor are detected by calculating the offset between the bounding box for the largest detected building and the center of the image.

Table 5: Difference in photo characteristics by correctly and incorrectly classified images

Machine	*Architects*						
	Georgian	Early Vic.	Late V./Edw.	Interwar	Postwar	Cont.	Revival
	Panel A: Total share of house blocked by vehicles or trees						
Georgian	–	0 (0.16)	-0.03 (-1.37)	-0.07 (-2.16)	-0.06 (-0.91)	-0.02 (-0.86)	0.02 (0.39)
Early Vic.	-0.04 (-2.80)	–	-0.04 (-7.42)	-0.11 (-8.26)	-0.09 (-5.45)	-0.03 (-1.97)	0.01 (0.47)
Late V./Edw.	0.03 (0.91)	0.03 (2.55)	–	0.01 (0.40)	0.01 (0.57)	0.01 (0.17)	0.08 (1.89)
Interwar	0.17 (2.24)	0.11 (3.71)	0.09 (5.73)	–	0.03 (4.16)	0.18 (4.42)	0.07 (2.40)
Postwar	0.13 (0.96)	0.03 (1.44)	0.11 (2.88)	-0.01 (-1.20)	–	0.09 (3.39)	0.04 (1.09)
Cont.	0.16 (0.77)	-0.02 (-1.91)	0.03 (1.14)	-0.03 (-0.92)	-0.04 (-3.06)	–	-0.03 (-1.32)
Revival	0.18 (0.74)	0.03 (0.92)	0.01 (0.26)	-0.05 (-3.03)	-0.06 (-4.90)	0.04 (2.02)	–
	Panel B: Share of house area						
Georgian	–	-0.11 (-3.43)	-0.10 (-2.67)	0.09 (1.94)	0.10 (0.96)	0.11 (1.29)	-0.02 (-0.31)
Early Vic.	0.05 (1.30)	–	0.03 (1.75)	0.07 (2.45)	0.18 (5.01)	0.12 (2.19)	0.05 (1.03)
Late V./Edw.	-0.09 (-2.08)	-0.07 (-3.10)	–	0.02 (1.07)	0.08 (2.23)	-0.08 (-1.29)	0.04 (0.71)
Interwar	0.08 (0.55)	-0.16 (-3.68)	-0.14 (-6.48)	–	0.03 (2.92)	-0.04 (-1.03)	-0.09 (-2.65)
Postwar	-0.33 (-1.98)	-0.24 (-3.82)	-0.18 (-3.25)	-0.05 (-4.60)	–	-0.12 (-2.78)	-0.16 (-4.57)
Cont.	-0.21 (-1.58)	-0.16 (-2.99)	-0.15 (-3.45)	-0.02 (-0.55)	0.01 (0.23)	–	-0.06 (-1.43)
Revival	-0.03 (-0.15)	-0.04 (-0.69)	-0.11 (-2.70)	0.06 (2.51)	0.08 (3.30)	0.05 (1.30)	–
	Panel C: Share of window area						
Georgian	–	-0.06 (-6.02)	-0.02 (-1.10)	-0.02 (-1.92)	0.03 (1.26)	-0.03 (-1.28)	-0.01 (-0.42)
Early Vic.	-0.01 (-0.60)	–	-0.01 (-1.51)	0.03 (2.72)	0.04 (5.30)	0.01 (0.70)	0.02 (1.83)
Late V./Edw.	0 (0.09)	-0.02 (-2.67)	–	0.04 (4.85)	0.02 (1.81)	0.02 (0.98)	0.02 (1.59)
Interwar	-0.03 (-0.92)	-0.09 (-10)	-0.08 (-13.20)	–	0.01 (2.98)	-0.03 (-2.41)	-0.03 (-3.48)
Postwar	-0.07 (-4.42)	-0.09 (-5.80)	-0.08 (-4.50)	-0.02 (-4.95)	–	-0.05 (-4.42)	-0.05 (-5.63)
Cont.	-0.07 (-2.04)	-0.08 (-6.97)	-0.05 (-3.73)	-0.01 (-1.46)	0.01 (2.50)	–	-0.02 (-2.58)
Revival	-0.02 (-0.18)	-0.06 (-4.39)	-0.03 (-1.71)	0.01 (1.04)	0.02 (2.50)	-0.02 (-2.33)	–
	Panel D: Image offset						
Georgian	–	0.04 (5.44)	0.02 (2.61)	-0.02 (-1.71)	-0.03 (-1.33)	0.01 (1)	0.01 (0.55)
Early Vic.	0.01 (0.64)	–	-0.02 (-6.59)	-0.03 (-3.82)	-0.04 (-4.98)	-0.01 (-0.93)	-0.03 (-2.78)
Late V./Edw.	0.04 (1.96)	0.03 (4.68)	–	-0.01 (-1.25)	-0.02 (-2.14)	0.01 (0.46)	0 (0.25)
Interwar	0.03 (1.16)	0.06 (6.28)	0.05 (9.75)	–	0.01 (2.35)	0 (0.06)	0.03 (3.19)
Postwar	0.04 (1.07)	0.06 (3.31)	0.04 (3.37)	0 (1.11)	–	0.02 (1.95)	0.06 (4.20)
Cont.	0.08 (2.11)	0.05 (3.87)	0.04 (3.40)	-0.01 (-1.32)	-0.02 (-4.26)	–	0 (0.20)
Revival	0 (0.02)	0.03 (2.06)	0.01 (1)	-0.01 (-1.85)	-0.01 (-1.02)	0.02 (2.22)	–

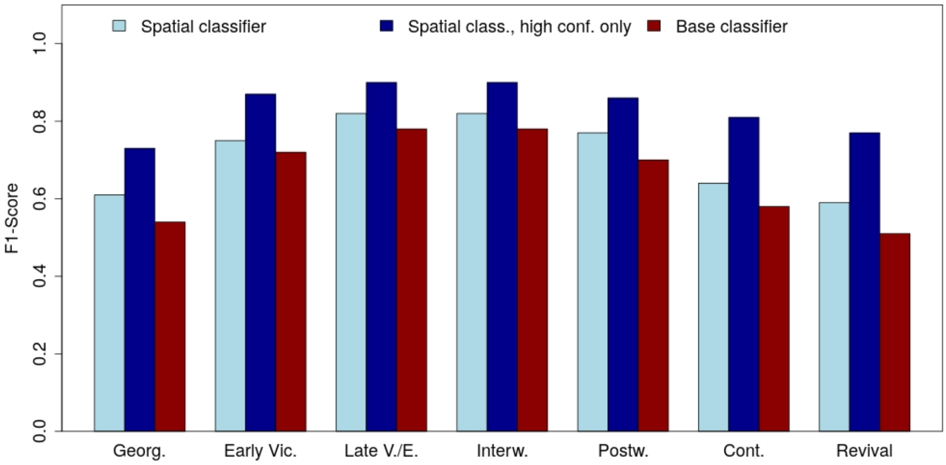
Notes: Cross-tabulation of photo characteristics by machine and architects' classifications. Off-diagonal elements ($Machine_j - Architects_i$) describe the difference in measure intensity between correctly and incorrectly classified photos for a given classification by architects. For example, cell {Early Vic., Georgian} compares the mean characteristic value for the set of photos which are actually Georgian but misclassified as Early Vic. to the set of correctly classified Georgian photos. T-stats are in parenthesis.

Figure 3: Examples of machine-based style classifications

Georg.	*Early Vic.*	*Late V./Edw.*	*Interwar*	*Postwar*	*Contemp.*	*Revival*

Notes: For a brief discussion of the building styles see page 6.

Figure 4: Difference of F_1-scores: spatial model vs. base model



Notes: The bar plot shows differences in the F_1-scores per category for the classifiers using both building level and neighbor information (dark and light blue) vs. a classifier using building level information only (red).

Table 6: Machine style classifications as dependent variables (Eq. 2 AIC comparison)

	No ML pred.	*Base pred.*	*Spatial pred., all*	*Spatial pred., high certainty*
Georgian	1,282	1,052	1,037	714
Early Victorian	10,231	7,597	7,259	4,594
Late Vic./Edw.	16,204	11,059	10,491	6,173
Interwar	18,333	14,471	13,637	7,670
Postwar	16,845	14,873	14,648	7,613
Contemporary	2,885	2,281	2,240	1,054
Revival	3,336	2,564	2,480	1,089

Notes: AIC values for 28 binomial models in which the occurrence of building styles (in rows) is explained by hedonic characteristics, location and time dummies (Column 1), augmented by the base ML predictor (Column 2), and the suggested spatial ML predictor (Column 3 & 4). Across building styles, the AIC are lowest for the spatial ML predictor, especially when uncertain predictions are omitted (Column 4). This suggests that the spatial predictor performs best.

Table 7: Hedonic Regression Estimates

	Dependent variable: *ln(price)*							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	13.20***	13.45***	13.29***	13.21***	15.61***	18.89***	13.14***	13.20***
	(0.20)	(0.17)	(0.20)	(1.74)	(1.72)	(3.77)	(0.20)	(0.20)
ln(dist. city center)	−0.45***	−0.50***	−0.47***	−0.48**	−0.72***	−1.12***	−0.44***	−0.45***
	(0.03)	(0.02)	(0.03)	(0.23)	(0.21)	(0.07)	(0.03)	(0.03)
Type: semi-detached	−0.12***	−0.14***	−0.13***	0.01	−0.12**	−0.10	−0.13***	−0.13***
	(0.01)	(0.01)	(0.01)	(0.04)	(0.05)	(0.07)	(0.01)	(0.01)
Type: terraced	−0.19***	−0.20***	−0.19***	−0.01	−0.10*	−0.12	−0.20***	−0.20***
	(0.01)	(0.01)	(0.01)	(0.04)	(0.05)	(0.11)	(0.01)	(0.01)
ln(area)	0.40***	0.40***	0.40***	0.33***	0.29***	0.21***	0.39***	0.39***
	(0.01)	(0.01)	(0.01)	(0.07)	(0.09)	(0.01)	(0.01)	(0.01)
ln(volume)	0.01***	0.01***	0.01***	0.01	0.02***	0.02	0.01***	0.01***
	(0.001)	(0.001)	(0.001)	(0.01)	(0.01)	(0.09)	(0.001)	(0.002)
New	0.09***	0.18***	0.13***				0.11***	0.10***
	(0.02)	(0.01)	(0.02)				(0.02)	(0.02)
	Base: Contemporary							
Georgian	−0.04	0.06***	−0.04				0.01	0.15**
	(0.03)	(0.02)	(0.03)				(0.03)	(0.06)
Early Vic.	−0.13***	−0.01	−0.11***				−0.04**	0.01
	(0.02)	(0.01)	(0.02)				(0.02)	(0.07)
Late Vic./Edw.	0.01	0.11***	0.04**				0.06***	0.15**
	(0.02)	(0.01)	(0.02)				(0.02)	(0.08)
Interwar	−0.15***	−0.02**	−0.13***				−0.08***	−0.15**
	(0.02)	(0.01)	(0.02)				(0.02)	(0.06)
Postwar	−0.22***	−0.06***	−0.19***				−0.12***	−0.17***
	(0.02)	(0.01)	(0.02)				(0.02)	(0.04)
Revival	0.04**	0.05***	0.03	−0.10	0.13***	0.12	0.05**	0.07*
	(0.02)	(0.01)	(0.02)	(0.09)	(0.04)	(0.09)	(0.02)	(0.04)
	Base: Neigh. Contemporary							
Neigh: Georgian							−0.10**	−0.44***
							(0.04)	(0.05)
Neigh: Early Vic.							−0.13***	−0.15***
							(0.02)	(0.04)
Neigh: Late V./Edw.							−0.05**	−0.15*
							(0.02)	(0.08)
Neigh: Interwar							−0.11***	−0.01
							(0.02)	(0.04)
Neigh: Postwar							−0.14***	−0.24***
							(0.02)	(0.04)
Neigh: Revival							−0.03	−0.03
							(0.02)	(0.06)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Neigh. dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interaction terms	No	No	No	No	No	No	No	Table 9
Observations	15,642	23,415	15,721	377	567	348	15,503	15,503
Adjusted R^2	0.88	0.88	0.88	0.93	0.91	0.92	0.89	0.89

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors are robust (White's estimator). Models (1) and (4) are estimated using the architects' classifications only. Models (2) and (5) are estimated based on all observations that have been automatically classified. Models (3) and (6) are also based on automatic classifications but the sample is reduced to high confidence predictions only. Models (4-6) are estimated on sales of newly completed buildings only. The interaction terms for Model (8) are in Table 9.

Table 8: Counts: Building style and neighboring buildings' style

Neigh.	*Building*						
	Georg.	*Early Vic.*	*Late V./Edw.*	*Interw.*	*Postw.*	*Cont.*	*Revival*
Georgian	164	44	7	0	0	0	1
Early Vic.	20	2205	274	34	18	18	27
Late Vic./Edw.	22	473	3128	226	22	9	10
Interwar	7	47	176	4165	450	22	45
Postwar	1	10	25	553	2712	38	32
Contemporary	0	0	5	16	42	348	12
Revival	2	17	5	62	36	19	153

Notes: The style of direct neighborhoods is defined as the most frequent detected style on the same street, within 100m.

Table 9: Coefficients interaction terms: Building and neighborhood style

Neigh.	*Building*						
	Georg.	*Early Vic.*	*Late V./Edw.*	*Interw.*	*Postw.*	*Cont.*	*Revival*
Georgian	0.19	0.25	0.56**	–	–	–	–
	(0.29)	(0.25)	(0.27)	–	–	–	–
Early Vic.	-0.14	-0.04	-0.13	0.03	0.07	–	0.10
	(0.18)	(0.09)	(0.12)	(0.09)	(0.08)	–	(0.10)
Late V./Edw.	-0.004	0.002	-0.01	0.18*	0.08	–	0.13
	(0.19)	(0.11)	(0.13)	(0.10)	(0.10)	–	(0.12)
Interwar	-0.23	-0.20**	-0.17	-0.03	-0.07	–	-0.15
	(0.19)	(0.09)	(0.11)	(0.08)	(0.06)	–	(0.09)
Postwar	0.18	0.03	-0.05	0.15**	0.14***	–	0.01
	(0.28)	(0.11)	(0.12)	(0.07)	(0.05)	–	(0.09)
Contemporary	–	–	–	–	–	–	–
	–	–	–	–	–	–	–
Revival	–	–	-0.43***	0.09	0.04	–	-0.04
	–	–	(0.15)	(0.08)	(0.07)	–	(0.09)

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors are robust (White's estimator). This table features the coefficients for interaction terms of building style and neighborhood styles only, while Table 7, Column 8 presents all other coefficients for this model.

Table 10: Combined effect: Sum of direct, neighborhood and interaction coefficients

Neigh.	*Building*						
	Georg.	*Early Vic.*	*Late V./Edw.*	*Interw.*	*Postw.*	*Cont.*	*Revival*
Georgian	-0.11	-0.17	0.27	–	–	-0.44	–
Early Vic.	-0.15	-0.18	-0.13	-0.28	-0.25	-0.15	0.02
Late Vic./Edw.	-0.00	-0.13	-0.00	-0.12	-0.24	-0.15	0.06
Interwar	-0.10	-0.19	-0.03	-0.20	-0.26	-0.01	-0.09
Postwar	0.09	-0.19	-0.13	-0.23	-0.27	-0.24	-0.16
Contemporary	0.15	0.01	0.15	-0.15	-0.17	0.00	0.07
Revival	–	–	-0.30	-0.09	-0.16	-0.03	0.00

Notes: The combined effect of building styles on property values is calculated by adding up the direct, neighborhood and interaction effects (Table 9). A revival building surrounded by other revival buildings, for instance, commands no premium over a contemporary building in a new neighborhood.

Bibliography

Ahlfeldt, Gabriel and Alexandra Mastro (Oct. 2012). "Valuing Iconic Design: Frank Lloyd Wright Architecture in Oak Park, Illinois". In: *Housing Studies* 27.8, pp. 1079–1099. ISSN: 0267-3037. DOI: 10.1080/02673037.2012.728575. URL: http://www.tandfonline.com/doi/abs/10.1080/02673037.2012.728575.

Asabere, Paul K., George Hachey, and Steven Grubaugh (1989). "Architecture , Historic Zoning , and the Value of Homes". In: *Journal of Real Estate Finance and Economics* 2.3, pp. 181–195. DOI: 10.1007/BF00152347.

Buitelaar, Edwin and Frans Schilder (2017). "The Economics of Style: Measuring the Price Effect of Neo-Traditional Architecture in Housing". In: *Real Estate Economics* 45.1, pp. 7–27. ISSN: 15406229. DOI: 10.1111/1540-6229.12137.

Coulson, N. Edward and Daniel P. McMillen (2008). "Estimating time, age and vintage effects in housing prices". In: *Journal of Housing Economics* 17.2, pp. 138–151. ISSN: 10511377. DOI: 10.1016/j.jhe.2008.03.002.

De Nadai, Marco et al. (Aug. 2016). "Are Safer Looking Neighborhoods More Lively? A Multimodal Investigation into Urban Life". In: arXiv: 1608.00462. URL: http://arxiv.org/abs/1608.00462.

Environment Agency (2015). *LIDAR Composite DSM - 1m.* URL: https://data.gov.uk/dataset/lidar-composite-dsm-1m1.

Francke, Marc K. and Alex M. van de Minne (Apr. 2017). "Land, Structure and Depreciation". In: *Real Estate Economics* 45.2, pp. 415–451. ISSN: 10808620. DOI: 10.1111/1540-6229.12146. URL: http://doi.wiley.com/10.1111/1540-6229.12146.

Fuerst, Franz, Patrick McAllister, and Claudia B Murray (2011). "Designer buildings: estimating the economic value of 'signature' architecture". In: *Environment and Planning A* 43.1, pp. 166–184. ISSN: 0308-518X. DOI: 10.1068/a43270. URL: http://www.envplan.com/abstract.cgi?id=a43270.

Gebru, Timnit et al. (2017). "Using Deep Learning and Google Street View to Estimate the Demographic Makeup of the US". In: *PNAS* 114.50, pp. 13108–13113. DOI: 10.1073/pnas.1700035114. arXiv: 1702.06683. URL: http://arxiv.org/abs/1702.06683.

Ghimire, B., J. Rogan, and J. Miller (2010). "Contextual land-cover classification: Incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic". In: *Remote Sensing Letters* 1.1, pp. 45–54. ISSN: 2150704X. DOI: 10.1080/01431160903252327.

Glaeser, Edward L., Michael Scott Kincaid, and Nikhil Naik (2018). "Computer Vision and Real Estate: Do Looks Matter and Do Incentives Determine Looks?" URL: http://www.nber.org/papers/w25174.

Glaeser, Edward L. et al. (Jan. 2018). "Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life". In: *Economic Inquiry* 56.1, pp. 114–137. ISSN: 00952583. DOI: 10.1111/ecin.12364. URL: http://doi.wiley.com/10.1111/ecin.12364.

He, Kaiming et al. (2016). “Deep Residual Learning for Image Recognition”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.

Helbich, Marco et al. (2013). “Boosting the predictive accuracy of urban hedonic house price models through airborne laser scanning”. In: *Computers, Environment and Urban Systems* 39, pp. 81–92. ISSN: 01989715. DOI: 10.1016/j.compenvurbsys.2013.01.001. URL: http://dx.doi.org/10.1016/j.compenvurbsys.2013.01.001.

Land Registry (2017). *Price Paid Data.* URL: http://landregistry.data.gov.uk/app/ppd.

Lindenthal, Thies (2017a). “Beauty in the Eye of the Home-Owner: Aesthetic Zoning and Residential Property Values”. In: *Real Estate Economics*, pp. 1–26. ISSN: 10808620. DOI: 10.1111/1540-6229.12204. URL: http://doi.wiley.com/10.1111/1540-6229.12204.

– (2017b). *Estimating Supply Elasticities for Residential Real Estate in the United Kingdom.* Comprehens. Elsevier Inc., pp. –. ISBN: 978-0-12-409548-9. DOI: https://doi.org/10.1016/B978-0-12-409548-9.09682-2. URL: http://www.sciencedirect.com/science/article/pii/B9780124095489096822.

Liu, Lun et al. (2017). “A machine learning-based method for the large-scale evaluation of the qualities of the urban environment”. In: *Computers, Environment and Urban Systems* 65, pp. 113–125. ISSN: 01989715. DOI: 10.1016/j.compenvurbsys.2017.06.003. URL: http://dx.doi.org/10.1016/j.compenvurbsys.2017.06.003.

Moorhouse, John C. and Margaret Supplee Smith (1994). “The Market for Residential Architecture: 19th century Row Houses in Boston’s South End”. In: *Journal of Urban Economics* 35.3, pp. 267–277. DOI: 10.1006/juec.1994.1016.

Naik, Nikhil, Ramesh Raskar, and César A. Hidalgo (May 2016). “Cities Are Physical Too: Using Computer Vision to Measure the Quality and Impact of Urban Appearance”. In: *American Economic Review* 106.5, pp. 128–132. ISSN: 0002-8282. DOI: 10.1257/aer.p20161030. URL: http://pubs.aeaweb.org/doi/10.1257/aer.p20161030.

Naik, Nikhil et al. (2017). “Computer vision uncovers predictors of physical urban change”. In: *Proceedings of the National Academy of Sciences* 114.29, pp. 7571–7576. ISSN: 0027-8424. DOI: 10.1073/pnas.1619003114. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1619003114.

Office for National Statistics (2017). *A Beginner’s Guide to UK Geography.* Tech. rep.

– (2019). *Open Geography Portal – 2011 Census Boundaries.* URL: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography.

Ordnance Survey (2017a). *AddressBase.* URL: https://www.ordnancesurvey.co.uk/business-and-government/help-and-support/products/addressbase-premium.html.

– (2017b). *OS Maps Local.* URL: https://www.ordnancesurvey.co.uk/business-and-government/products/os-open-map-local.html.

Plaut, Steven and Egita Uzulena (2006). “Architectural Design and the Value of Housing in Riga, Latvia”. In: *International Real Estate Review* 9.1, pp. 112–131.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin (2016). “"Why Should I Trust You?" Explaining the Predictions of Any Classifier”. In: *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, pp. 1135–1144.

Szegedy, Christian et al. (2015). “Rethinking the Inception Architecture for Computer Vision”. In: ISSN: 08866236. DOI: 10.1109/CVPR.2016.308. arXiv: 1512.00567. URL: http://arxiv.org/abs/1512.00567.

The Economist (Nov. 2018). “The line of beauty”. In: *The Economist*. URL: https://www.economist.com/britain/2018/11/17/how-to-defeat-nimbyism-build-more-beautiful-houses.

Vandell, Kerry D. and Jonathan S. Lane (1989). “The Economics of Architecture and Urban Design: Some Preliminary Findings”. In: *AREUEA Journal* 17.2, pp. 235–260.