Machine Learning, Architectural Styles and Property Values

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

14 December 2020

Abstract

This paper couples a traditional hedonic model with architectural style classifications from human experts and machine learning (ML) enabled classifiers to estimate sales price premia over architectural styles, both at the building and the neighborhood-level. We find statistically and economically significant price differences for houses from distinct architectural styles across an array of specifications and modeling assumptions. Comparisons between classifications from ML models and human experts illustrate the conditions under which ML classifiers may perform at least as reliable as human experts in mass appraisal models. Hedonic estimates illustrate that the impact of architectural style on price is attenuated by properties with less well-defined styles and we find no evidence for differential price effects of *Revival* or *Contemporary* architecture for new construction.

Introduction

In this paper, we introduce a new data collection technique that allows us to rigorously test the notion that households have strong preferences over architectural styles. Our results indicate that 1) there is evidence for sales price premia associated with houses from a variety of architectural styles, 2) hedonic estimates from high confidence machine learning (ML) enabled architectural image classifications are similar to human expert estimates, and 3) any style premia are found for existing structures only and not for new buildings, indicating that debates about strong demand for historicizing aesthetics may be unfounded.

Understanding household preferences for architectural styles is of increasing interest to both policymakers and researchers. The policy perspective is best illustrated by Britain's *Building Better, Building Beautiful Commission*, which advises the government on design choices for homes and neighborhoods. It lists as 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) and boldly claims that matching the style of new housing to aggregate neighborhood preferences should overcome otherwise prevailing objections of incumbent households to new construction.

Researchers have recently begun to quantify and define what makes for *better design and style.* Buitelaar and Schilder (2017) find evidence on the link between preferences over architectural styles and housing prices using architectural assessments by human experts, and 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. 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 vintage effects. They find that buildings from the 1930s carry a strong price premium. These results support the hypothesis that preferences for house vintages (which coincide with architectural styles) may affect housing prices.

Clearly, building vintage, quality, location, and architectural styles are highly correlated which implies that hedonic price estimates for architectural styles cannot simply be understood as marginal prices for aesthetics. We only interpret our empirical findings as 'pure' architectural preferences in few cases where differences in all other dimensions can be accounted for, e.g. new units at similar locations that have been built to similar standards and that have not depreciated yet. For markets with more contemporary housing stocks, differences in functional obsolescence would be a major concern when age and style variables are correlated. In our case, over 93% of the housing stock had been produced prior to 1980 and

70% before 1940, so concerns on contemporaneous depreciation of new structures may be alleviated as the vast majority of homes has already depreciated. Nevertheless, systematic differences in locations and unobservable quality characteristics and modernization standards cannot be ruled out.

Buitelaar and Schilder (2017) indicate that any premium for an architectural style must stem from either differences in construction prices or from supply constraints, as new construction potentially does not capture the demand for traditional styles. In markets where home builders are free to decide on the architectural style of the homes they supply, we would not expect any price effects beyond differences in construction costs. If there was, for instance, a premium for more traditional styles, developers would continue to build more of these until all demand is met. Our prior is therefore *not* to find distinct price effects of aesthetics for newly supplied homes.

Of course, residential buildings rarely 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. A similarly shaped neighboring building is value-enhancing while proximity to a wildly different neighboring shape, everything else remaining equal, may be detrimental to property values (Lindenthal 2020).

In traditional approaches chosen by e.g. Buitelaar and Schilder (2017) or Ahlfeldt and Mastro (2012), each observation is classified into an architectural style by a human expert, which is time-consuming, costly and is not feasible for large sample sizes. As Helbich et al. (2013) state, 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. In addition, an expert's views may not fully reflect the market's perceptions of relevant styles – architectural classifications and economically relevant segments might differ.

Combining automated Google Street View photo data collection with deep learning/ML image classification offers a promising way forward to large and comprehensive data sets of architectural style. Naik et al. (2017) describe 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 to the block, street, or street-section level classifications used in the majority of previous studies that use automated image classification, Glaeser, Kincaid, and Naik (2018) push the level of observation 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 increasingly accurate and objective manner and may be implemented in a more efficient, accessible, and cost-effective way.

In addition to reducing the cost of data collection, the ability for low-cost classification of a large sample of structures in an urban area enables researchers to detect not only the architectural style of the building but also to characterize the style of other buildings in the vicinity. This allows us to examine the interaction between the style of a building and that of its neighbors.

To analyze the accuracy of our method of detecting architectural style, we compare our algorithmic results with an extensive data set of architectural style classifications compiled by human experts. The relatively large and costly comparison group provides insights into ML-based classification accuracy and the robustness of hedonic estimates to human and ML classifications. In addition, the measurable uncertainty in the ML classifications may also be a useful source of variation to exploit for style identification. For example, it may be the case that the impact of architectural style on sales price for existing buildings may be attenuated for buildings that are more difficult to classify. To our knowledge, this step is not present in the existing literature and provides additional insight into the efficacy of the technique for application in other domains.

We find evidence that for existing buildings, the architectural style has an impact on the sales price and that the estimates are very similar for both human expert and high confidence ML classified images. The price effects of human-expert defined architectural styles are greater in images where there is agreement with the ML classifications. The presence of architectural style in our hedonic specifications reduces the RMSE of our models. The preferred architectural style for resales appears to be *Revival* with *Interwar*, *Postwar*, and *Early Victorian* commanding the smallest premium. Immediately proximate neighbor styles

---

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

also have an impact on sales price, with Contemporary neighborhoods clearly preferred to Georgian, Early Victorian, and Postwar. Importantly for policymakers, we do not identify a premium for either Revival or Contemporary styles for new construction.

We begin by describing the traditional hedonic data used in the paper. Second, we describe the methodology for the automated architectural style classification. Third, we compare the results of the automated predictions with expert classifications. Next, we present an array of hedonic regressions using both architect and ML-based classification and discuss results. Finally, we conclude with a discussion of the policy implications of the estimates and the feasibility of ML-based classification of architectural style.

Data

Residential real estate transactions in England and Wales are collected and published by the Land Registry (Land Registry 2017). Their records include the date of transaction, the 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, England, which were recorded between January 1995 and October 2018, excluding any leaseholds, apartments and properties classified as type "other", and sales with prices below £50,000 or in excess of £2,000,000. Notably, the Land Registry data lack relevant variables such as year of construction, home and lot size, or similar quality indicators. We, therefore, augment the transaction data with core hedonic variables from other sources and to control for a variety of building-specific effects that may be correlated with both sales price and architectural style.

The Ordnance Survey (OS)[2] provides high-resolution maps that show the two-dimensional footprint of buildings in the UK. Using these maps, we calculate each building's floor plate (in m^2) and estimate its volume by combining the building outlines with digital elevation models from the Environment Agency (2015), as suggested by Lindenthal (2017). Additionally, we measure the distance from each residential property to the city center, proxied by Great St. Mary's Church. The Office for National Statistics (2019) subdivides Cambridge into 69 unique *Lower Super Output Areas* (LSOA) and we rely on these boundaries when constructing neighborhood dummy variables. An LSOA typically has 1,000–3,000 residents and 400–1,200 households of comparable economic and socio-demographic characteristics (Office for National Statistics 2017).

Architectural Style Data

Our main variable of interest, architectural style, follows the classifications commonly used by realtors, home-buyers, and architects. Styles are distinctive in Cambridge: Table 1 provides a description of the

---

[2]https://www.ordnancesurvey.co.uk/

construction eras and the distinctive motifs for each style as defined by members of the Architecture Department at the University of Cambridge.

– Insert Table 1 about here –

We aim to collect images of all residential buildings in Cambridge from Street View. Approximately 50,000 buildings found on the OS maps have footprints between 30 and 500 m^2. For 85% of these candidates, we can download street-level images on Google Street View. However, not all of them turn out to be residential buildings. A sample of 25,000 of these images is categorized by final year architecture students into the seven styles.

We can match 23,768 residential properties sold in Cambridge between 1995 and 2019 to images from Google Street View. We implicitly assume that houses have not changed their architectural styles between the time of sale and the time Google collected the images. This is a reasonable approach since conversions of just the exterior of buildings rarely occur. In case of full redevelopments, we exclude any sales for that address pre-dating the redevelopment, since we cannot rule out a change of style at redevelopment.

Importantly, 15,511 sales can also be matched to the sample of expert's classifications. Our hedonic models will be estimated on these observations for which we have predicted and ground truth classifications. This allows us to assess the prediction accuracy and robustness of the regression results to the classification approach. When assessing the architectural style of surrounding buildings, we use the full set of all images collected. Being able to cover almost the universe of all buildings in a market is a clear advantage of an automated approach.

In Figure 1, we present a set of representative images for each of the 7 architectural styles. These images come from our automated collection procedure and are a visual representation corresponding to the expert descriptions in Table 1. Image information here is very useful as it provides a visual mapping from the style characteristics to actual identification. One component which is evident in the images but not necessarily in the descriptions is the within style homogeneity of window grouping, relative pane size, and counts. For example, Georgian windows contain 12 equally sized panes in 4 rows and 3 columns, Early Victorian windows consist of 4 equally sized panes in 2 rows and 2 columns, Interwar windows contain an upper smaller pane and larger lower pane. Interestingly, the Revival style demonstrates heterogeneous window patterns that borrow from the variety of classic architectural styles. We revisit the impact of image quality and window visibility on prediction accuracy below.

– Insert Figure 1 about here –

Image collection

To provide the best information to detect style for both the architect and ML classifiers, images should be focused on the structure under examination, have a clear field of view, show minimal picture overlap with neighboring properties, and minimize potentially confounding issues such as vehicle type.

Unfortunately, in the UK, Google's Street View API regularly fails to identify the building at a given address. Instead, the camera location will be in close proximity to the front of the building but aimed straight ahead and down the center of the road. Fig. 2 (a) presents a typical result from an address level API. Here, the Google Street View API algorithm fails to accurately capture the front of the building and the property is not even contained in the photo. For major US cities, the accuracy of Google's search results is higher and the Street View camera will pan towards the center of the property parcel. This makes the photo collection of studies such in U.S urban areas potentially much more straightforward although the API may result in images that are too 'zoomed out' resulting in images that may contain obstructed views, vehicles, and neighboring properties.

– Insert Figure 2 about here –

Our automated image collection algorithm utilizes a sophisticated process of viewshed analysis based on Ordnance Survey building maps and ML methods to capture images using the best camera locations, zoom, and angles while accounting for potential obstructions from buildings and trees or vehicles. Thus, for each property, our goal is to automate the collection of the best possible Street View image for each property in our sample.

More specifically, we first use OS maps to identify each building's outline (see Figure 3). Second, we build a database of all possible Google Street View camera locations using Metadata queries of addresses in the Ordnance Survey. These metadata queries return the latitude, longitude, and dates of the most recent panorama photos collected by Google Street View. Next, we use the OS data to assign each property to the nearest Street View photo location in the metadata database and conduct a viewshed analysis on surrounding building outlines to assign pan and zoom parameters to find an informative view. The algorithm allows us to estimate the camera bearing (green line) and zoom factor, based on the fan of the lines of sight (in blue). The line of sight criteria identifies which exterior walls are visible from the nearest Street View point and aims away from any wall segments where the direct line of sight is obstructed by other buildings. A visual example of this algorithm is shown in Figure 3.[3]. For each property, we then collect the image at the highest resolution offered by Google's Street View API[4] (640x640 color pixels). We test the robustness of our hedonic models to a variety of image quality concerns in the results section of the paper.

---

[3]Code necessary to replicate the full image capture including metadata capture and viewshed based camera aiming is available from https://github.com/thies/paper-cambridge-vintages

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

7

Figure 2 (b) shows how the suggested approach can improve the quality of the harvested images. The new image is focused appropriately on the relevant property, it contains minimal neighbor information. and the view is not overly obstructed by other structures, greenery, or vehicles. As a result, the image contains the relevant identifying architectural style characteristics such as a clearly identified roofline, front door, window panes, and bay window.

Methodology: Style classification

With the images collected, we next focus on training an ML learning model to classify the architectural style observed in each of the collected Street View images. The first step in building our ML model is to construct a training data set. Our training data set consists of images from 25,000 randomly sample properties classified into one of seven different architectural styles by final year students from the Architecture Department at the University of Cambridge[5] A sample of 25,000 is much more training data than is actually needed. By applying transfer learning to an existing image recognition model, we require less than 250 observations per category to reach saturated training accuracy levels. We use this "surplus" training data to perform out of sample comparisons between our ML predictions and human expert classifications. This large set of expert classified data also allows us to perform robustness, and falsification checks on the hedonic regressions for the 15,369 overlapping observations where we have transaction data with both human and ML-based style classifications.

Training the model

We simplify the calibration of our architectural style prediction model by utilizing transfer learning. Transfer learning freezes the parameters from a pre-trained image recognition model, removes the final steps in the model, and then trains a new model on the vector of outputs from the truncated pre-trained model to produce new classifications[6]. Transfer learning methods enable the use of smaller training data sets and impose a much lower computational burden than traditional deep learning models. 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)[7], which evaluates image classification and object detection algorithms for a wide range of objects. The pre-trained classifications would allow us to identify pets, vehicles, or people in the pictures – assessing architectural style, however, is beyond the canned classifiers' capabilities. To allow for the detection of new objects, transfer learning techniques strip off the last steps of the model. The output of the original deep learning

---

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

[6]For examples, see https://www.tensorflow.org/tutorials/images/transfer_learning

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

model is essentially just a deeply recursive dimension reduction engine that outputs a 2048-dimensional feature vector for each picture. This feature vector describes the outline shapes, locations, and colors that were important for the ImageNet classifications.

Glaeser, Kincaid, and Naik (2018) rely on a 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). In our analysis, we follow a different strategy and actually double the dimensionality (4096) of our feature vectors 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).

Having obtained feature vectors for each image, we train a simple multinomial classifier comprising of 1) an input layer the size of the feature vectors, e.g. 2048 or 4096, respectively, 2) one dense layer rectified linear activation function (relu) half the size of the input layer, one subsequent dropout (rate 0.5) layer, 3) one dense layer (relu) a quarter the size of the input layer, 4) one subsequent dropout (rate 0.5) layer, and 5) the final dense output layer with softmax activation. The dropout layers help avoid overfitting, thus increasing the generalizability of the predictors. 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.[8] All classifiers are implemented using the Keras/TensorflowAPIs[9]. The computational burden of this rather shallow model design is modest.

Next, we turn to an analysis of the predictive accuracy of the models under a variety of assumptions and diagnostic tests.

Classification Diagnostics

We use a variety of methods to quantify and explain the variation in ML prediction accuracy under differing modeling assumptions. First, we report the accuracy of the predictions using a confusion matrix (a contingency table of true and predicted classifications) that also offers recall, precision, and F_1 scores. Second, we explore how the style classification prediction error is affected by the modeling assumptions. Third, we test the sensitivity of predictions to variation in the training data. After quantifying the behavior of the style predictive models, we next turn to describe how image characteristics such as aim quality affect the accuracy of the classifier. This final analysis should help inform other researchers and practitioners in best practices for image collection in order to maximize predictive accuracy.

---

[8]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 ML Classification Diagnostics section.

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

Electronic copy available at: https://ssrn.com/abstract=3604052

Table 2 contains confusion matrices for two different ML classifiers and a 'high confidence' prediction criteria. Each confusion matrix measures how the ML style predictions compare to the architects'. Panel A illustrates the predictive performance for an ML model based on an image of both a building and its nearest neighbor. Panel B includes only the highest confidence predictions from the spatial classifier used to construct Panel A. Panel C contains predictions from an ML classifier based on a single building image.

Diagonal elements in the confusion matrices represent the count of images where the ML and the architect agree while off-diagonal elements are cases where the ML misclassifies the style. Consider the *Interwar* style in Panel A. There are 5,884 images where the ML correctly classifies the style. For the *Interwar* column, the ML incorrectly classifies 38 images as *Georgian*, 86 as *Early Victorian*, 213 as *Late Victorian/Edwardian*, 514 as *Postwar*, 101 as *Contemporary*, and 139 as *Revival*. Thus, of 6,975 true *Interwar* images, the ML misclassifies 1,091 and correctly detects 5,884. Recall measures the ability of the ML to detect within a given category and is the share of correctly classified images out of the total for a category. For *Interwar* the recall rate is then 5,884/6,975=0.84.

Now consider the *Interwar* row which contains information on all images that the ML classifies as *Interwar*. There are 10 *Georgians*, 46 *Early Victorian*, 254 *Late Victorian/Edwardian*, 997 *Postwar*, 56 *Contemporary*, and 54 *Revival* incorrectly classified as *Interwar*. In total, there are 1,417 images incorrectly classified as *Interwar* and 5,884 correctly classified. Thus, 81% of the images classified by the ML as *Interwar* are actually *Interwar*. This is the Precision or the share of buildings predicted to belong to a category that is actually from the category. The $F_1 - score$ is the harmonic mean of Precision and Recall.

Across models, classification error is most likely to occur in styles that occurred in consecutive eras. For example, in Panel A, 10% of *Late Victorian/Edwardian* styles are classified as *Early Victorian* and 14% of *Georgians* are misclassified as *Earl Victorian*. On the other hand, for distant eras, the misclassification likelihood is much lower and the ML exhibits less confusion. The more classical *Georgian* and *Early Victorian* styles are each only assigned to Interwar 3% of the time.

Recall, precision scores, and F_1 scores are higher for the spatial ML classifier shown in Panel A relative to spatially naive ML classifier in Panel C. Pronounced increases include the increase in recall for *Postwar* from 0.61 to 0.71 due to a reduction in *Postwar* images being incorrectly classified as *Interwar*. Also notable is the increase in recall for *Revival* from 0.58 to 0.67.

Table 2 Panel B is based on the spatially dependent model but only contains classifications where the maximum ML style prediction score was above 0.8. These 'high confidence' ML images improve the measurable performance of the ML model relative to the full sample. Across styles, F_1 scores see a substantial increase. Notably, recall rates for *Early Victorian* and *Interwar* are now above .90 and *Contemporary* increases from 0.72 to 0.86. In other words, images that are easier for the ML to classify

are also the ones where the ML is correct at a higher frequency.

Figure 4 shows the distribution of F_1 scores from Table 2 across ML modeling assumptions and architectural styles. This figure demonstrates the benefits of including the neighbor image in our modeling strategy. In addition, the marginal contribution to the F_1 score from using a spatial classifier appears to be of a similar magnitude to that of including only 'high confidence' images and greatly improves model accuracy without the cost of dropping images from the sample.

<center>– Insert Table 2 and Figure 4 about here –</center>

Next, we regress the architects' style classifications captured by binomial variables $D_{Architects}$, on a vector of hedonic variables $\mathbf{X}$, vectors of year $\mathbf{Y}$ and neighborhood $\boldsymbol{Loc}$ dummy variables, and the automatically estimated style $\mathbf{ML}$. 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. These models are estimated both with and without spatial dependencies and capture the ability of the ML to predict the architect style classifications net of other hedonic variables.

$$logit(D_{Architects,i}) = \alpha + \beta\mathbf{X_i} + \delta\mathbf{ML}_i + \gamma\mathbf{Y}_i + \lambda\mathbf{Loc}_i + \epsilon_i \tag{1}$$

The logit regressions are estimated by generalized least squares. For a well-performing classifier, the δ coefficients will be statistically significant and, more importantly, the Akaike Information Criterion (AIC) will be relatively low. A summary of AIC results is shown in Table 3. Column 1 (*No ML pred.*) contains the AIC for style prediction models based solely on the hedonic variables. Column 2 (*Base pred.*) includes both the hedonic variables and the single image (spatially naive) ML. The AIC for all styles decreases with the inclusion of this 'base' ML predictor. In Column 3 (*Spatial pred. all*), we include hedonic variables along with the spatially aware ML predictor. Again, the ability to predict improves across all architectural styles. The improvement is especially notable in the final column (*Spatial pred. high certainty*), where the AIC is cut in half for the 'high certainty' subsample of the spatial predictors. As expected, the AIC rankings are consistent with F_1 rankings in Table 2. By including the hedonic variables as well as location and time dummies in the regression, we are able to see that building style is not fully explained by size, volume, location, and building type. The AIC results clearly show that 1) a model based on hedonic variables alone does not predict actual style as well as a model that includes the predicted style from the image and 2) subsetting on high confidence ML classified images increases AIC.

<center>– Insert Table 3 about here –</center>

Subsequently, we measure the sensitivity of our predictions to sampling variation in the training data. This test is only possible due to the extremely high number of training data images. We begin by creating

<center>11</center>

100 models that are calibrated using 250 randomly selected training data images for each style category in the data set. For each image in the out-of-sample data, we then capture the highest probability style type prediction. With these predictions, we perform two calculations. First, we utilize these predictions as an ensemble model and assign the most commonly predicted ML class as the architectural style. For example, if an image is predicted to be a Georgian architectural style 80 of 100 times (80 "votes"), Early Victorian 10 times, Late Victorian 5 times, and Revival 5 times the building is assigned the Georgian style. Second, we use the predictions to generate a Herfindahl index for each image. In the example case the Herfindahl index for image i is calculated as

$$Herf_i = \sum_{s=1}^{7} \left(\frac{votes_s}{votes_{all}} \right)^2 = 0.8^2 + 0.1^2 + 0.05^2 + 0.05^2 = 0.655 \tag{2}$$

The Herfindahl scores for misclassified images (off-diagonal elements in Table 4) indicate are all negative. This provides strong evidence that high consensus ensemble models are more likely to be accurate than low consensus models. All off-diagonal elements in the lower panel of Table 4 are negative, suggesting that, in general, the Herfindahl index is a good predictor of correctly classified images. In other words, there is less consensus for misclassified images.

The next step examines the *objects* in each image. This serves two purposes 1) we can gain a better understanding of how image composition affects our ML prediction accuracy and 2) we can validate the quality of our automated image collection algorithm. We again turn to ML automation and use an off-the-shelf object detection algorithm (Inception/Resnet) that can identify broadly defined objects in images such as trees, vehicles, houses, doors, or windows without any additional training.[10] With this information, we are able to validate the performance of the viewshed collection algorithm according to several measures.

First, we identify the locations (bounding boxes) of cars, trees, buildings, and windows in each image. Next, we construct a variety of measures of image quality including house area, share blocked, window area and image offset.

We measure "house area" as the share of the image that is taken up by the house bounding box.[11] Share blocked measures how obstructed the view of the house is by cars or trees. Recall from the viewshed exercise that we aim away from structures in order to get a clean line of sight. However, buildings may have obstructed lines of sight due to greenery, fences, garden walls, or large vehicles that cannot be detected from the Ordnance Survey maps. A high 'shared blocked' image should provide less information for our ML style classifier and result in lower accuracy. As discussed in the architectural style descriptions, one of

---

[10]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

[11]If more than one house is detected, then the largest of the houses is used as the measure.

the most important visual differences across architectural styles is window alignment and composition. We quantify this with the measure 'window area' which calculates the share of the image area that contains windows. We use "image offset" to detect images not collected at an optimal angle or zoom factor. To quantify this, we calculate the offset between the bounding box for the largest detected building and the center of the image. [12] Table 5 shows the mean values by architectural style for each of the quality measures. Overall, *Georgian* and *Contemporary* styles have the highest quality scores. Even for *Postwar*, our lowest-performing style, the average image contains 64% house with windows taking up 7% of the image.

We explore the relationship between the image quality measures classification accuracy in Table 6. Panel A compares the mean values for views obstructed by trees or vehicles for correctly (on-diagonal) and incorrectly (off-diagonal) classified images. In most cases, the off-diagonal values are positive as obstacles in the line of sight correlate with a higher likelihood of misclassifications. This helps alleviate concerns that the style detection model is loading on trees, cars, or other building structures. This also shows that ML misclassification is more likely to occur when the building is less directly observable. If the house takes up a larger share of the image (Panel B), misclassifications are reduced (negative differences). The strongest effect is for the window area, which appears to give valuable cues about a house's style (Panel C). Finally, images with houses that are not in the center of the frame tend to be misclassified more frequently (positive differences in Panel D). In summary, image quality appears to be correlated with style prediction accuracy.

<center>– Insert Table 5 and Table 6 about here</center>

Summary statistics for the transaction level dataset are shown in Table 7. New construction is only 4% of the sample and terraced and semi-detached residential properties comprise 87% of the sample. Importantly, based upon architectural style, the residential housing stock in Cambridge is old with only about 7% of properties from post-1980 eras of Contemporary and Revival architecture. Interestingly, the two most common architectural styles are the ubiquitous 20th-century styles of Interwar and Postwar which together make up 55% of all sales. Classical styles such as Georgian, Early Victorian, and Late Victorian/Edwardian together account for 39% of sales. Neighborhood style shares seem to closely match the individual style shares which hint at a high degree of architectural style clustering in the city.

<center>– Insert Table 7 about here –</center>

Figure 5 visualizes the spatial distribution of the regression sample. The maps trace Cambridge's historic expansion, with *Georgian* houses clustering in the city center, surrounded by first *Early Victorian* then *Late Victorian/Edwardian*, *Interwar* and *Postwar* homes. *Contemporary* and *Revival*, however, are widely

---

[12]Higher offset scores may result from either imperfectly defined ordnance building outline data (which may occur with some frequency), another building obstructing the viewshed algorithm causing the camera to pan or zoom away from the obstacle, or a tree or car causing a complete view obstruction in which case the algorithm will move to the next best location.

<center>13</center>

distributed across Cambridge's neighborhoods, usually filling in previously undeveloped or brownfield sites, densifying the city. Neighborhoods might be dominated by a single building style - but only very few comprise of more than 80% of buildings of one style. These extreme cases are only found in large *Interwar* or *Postwar* expansion areas. Luckily, many more houses from these frequent styles are found in other neighborhoods, which avoiding excessive multicollinearity between style and neighborhood variables.

– Insert Figure 5 about here –

Hedonic Estimation

The hedonic estimations use both the ML and architect classifications for a variety of direct, robustness, and falsification tests. Merging the ML classifications 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:

$$ln(Price_i) = \alpha + \beta\mathbf{X_i} + \delta\mathbf{Style}_i + \eta\mathbf{StyleNeigh}_i + \iota\mathbf{Style}\cdot\mathbf{StyleNeigh}_i + \gamma\mathbf{Y}_i + \lambda\mathbf{Loc}_i + \epsilon_i \quad (3)$$

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 $\mathbf{Loc}$ dummy variables and the building's estimated $\mathbf{Style}$ and the prevailing styles of other buildings in the direct proximity ($\mathbf{StyleNeigh}$). $\mathbf{Style}\cdot\mathbf{StyleNeigh}$ 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? For Cambridge, 47% of new supply has been of contemporary and 36% of or revival style. Estimating Eq. 3 for a subset of newly constructed buildings from these two styles 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 and quality of these buildings, which is tightly intertwined with their aesthetics.

Results

The estimated coefficients from 8 different versions of the hedonic regression specified in Eq.(3) are reported in Tables 8 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 of sale 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.

<div align="center">– Insert Table 8 about here –</div>

Column (1) in Table 8 presents the estimated regression coefficients for a rudimentary base model: Hedonic variables, location and time dummies jointly explain more than three-quarters of the variation in transaction prices and the root mean square error (RMSE) is 0.235.

Adding the architects' classifications as explanatory variables in Model (2), reduces the RMSE to 0.225 - which is 4 percent lower than the error in (1). The coefficients on the architectural styles reveal an interesting and intuitive pattern: 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: *Postwar* exhibits the strongest price discount (-0.22), followed by *Interwar* (-0.14), and *Early Victorian* (-0.13). These discounts are economically and statistically significant. We do not detect a significant price differential for *Georgian* and *Late Victorian/Edwardian* properties, while *Revival* architecture appears to command a small premium ($+0.04$) over homes built in the *Contemporary* style. Model (3) uses the ML predictions for the style variables instead of the architect's expert classifications. Overall, the style coefficients become more positive relative to the base. It must be the case that there is a negative correlation between properties incorrectly classified as *Contemporary* and sales price. This may result from the disproportionate misclassification of more attractive or higher quality *Contemporary* to one of the other styles or less attractive buildings being misclassified as *Contemporary*. This bias vanishes in Model (4), which uses only ML classifications of high confidence (at the cost of dropping 1/3 of the sample). The RMSE creeps slightly up for Models (3) - but is lower for (4) than for the model with the architect defined style covariates (admittedly, the sample has changed, so the comparison is not fair). Importantly, Model (4) shows that the ML-based results approximate the architects' estimates (Model 2) if the analysis is restricted to high confidence classifications.

Model (5) is a falsification test to investigate if unobservable variables correlated with style are driving our results. We select all cases in which the experts and the ML classifier disagree (3,063 observations) and re-estimate Model (2). This basically leaves only the 'difficult to classify' images. If the correlation of architectural style with unobserved building quality, age, interior updates, backyard sizes, etc. are actually driving results then these estimates should be the same as those in Model (2). If, on the other hand, distinct and easily recognizable architectural style really matters and correlated unobservables are not driving the results, then these point estimates should be insignificant. Interestingly, the architects'

(true!) classifications cease to have a strong price effect. All coefficients gravitate towards 0 – even the significant discount on *Postwar* is cut in half. Apparently, if exterior styles are more difficult to categorize, the style classifications might still matter to architects but are less meaningful for buyers and sellers. For this reason, we base the remaining analysis on the set of high confidence machine observations where the observable hedonic attributes of architectural style are more clearly defined.

In Model (6), we control for age and quality by selecting only transactions of new homes of either *Contemporary* or *Revival* styles.[13] Basically, this contrasts houses that might *look* different from the street, having either contemporary or revival façades, but which are all modern homes at their core. All have been built according to modern specifications, using the same technologies and materials, and only differ in their appearance. After controlling for location, building characteristics and quality, buyers show no willingness to pay a premium for *Revival* architecture.

Exploring the impact of the prominent style of buildings on the same street and within 100 m (**StyleNeigh**) of the sales property, we find that the magnitude and significance of the own style estimates in Table 8, Column 3 and Column 7, are stable and the rank ordering of the style impact is maintained. Neighbor style does have a small if muted impact on prices. *Postwar*, *Early Victorian* and *Interwar* buildings are the least popular neighbors. Other neighbors are economically and statistically insignificant relative to a *Contemporary* neighborhood. Importantly, we find no evidence in favor of positive externalities from revival architecture: the difference between *Revival* and *Contemporary* neighbors is not statistically significant. Everything else equal, buildings within ensembles of *Revival* do not achieve higher transaction prices than buildings in more historic areas (as Table 11 will confirm later).

Table 9 presents the ι interaction coefficients (Eq. 3) for style-neighboring style combinations. Some style combinations have insufficient numbers of observations, as tabulated in Table 10, and drop out. Neighboring styles are compared to the base case of a *Contemporary* house in a *Contemporary* street. We find that after controlling for neighborhood and building style, there is little additional impact from the building neighborhood interaction terms. The main exception to this rule is the positive impact that a *Georgian* style building in a *Georgian*, *Early Victorian*, or *Late Victorian/Edwardian* neighborhood. The negative (although statistically insignificant) coefficients on dissimilar neighboring styles support a finding by Lindenthal (2017) that shows that a harmonious match of a building's shape with its direct environment may lead to a price premium.

– Insert Table 10, Table 9 and Table 11 about here –

In Table 11, the combined effect of building styles on property values is calculated by adding up the direct, neighborhood and interaction effects (Tables 8 and 9). When filling in lots in historical neighborhoods,

---

[13]The comparison may only be done on these style types as the buildings with the other architectural styles are inherited from previous eras and are no longer constructed.

16

however, a premium for *Revival* facades over modern designs can be observed (differences in the last two columns of Table 11). For large scale new construction, however, *Revival* buildings surrounded by other historical buildings do not sell at a premium relative to *Contemporary* buildings in a more modern setting.

As a final robustness check, we present rerun the pricing models and include with year and neighborhood ($\mathbf{Y} \times \mathbf{Loc}$) interaction effects. The main limitation of introducing the interaction is that we do not have enough observations to run Model (5) and which was based on the set of new constructions due to the limited sample size. The results are shown in Table 12.

– Insert Table 12 about here –

Conclusion

In this paper, we study the relationship between architectural style and residential sales price. We present evidence for economically significant price differences between buildings from different architectural styles – but not for architectural style on its own. Analyzing house transaction prices from Cambridge, we find that any marginal price estimates are very similar for classifications by both human experts and confident ML classifiers. The preferred architectural style for resales appears to be *Revival* with *Interwar*, *Postwar*, and *Early Victorian* commanding the smallest premium. Immediately proximate neighbor styles also have an impact on sales price, with Contemporary neighborhoods clearly preferred to *Georgian, Early Victorian,* and *Postwar.* The effects are much more strong in images where architectural style is easily defined.

Importantly, for policymakers, we do not identify a premium on either *Revival* or *Contemporary* styles for new construction. After accounting for building quality and location, which tend to be better for older vintages, no evidence for a premium for historicizing architecture 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.

Our analysis also highlights that it is very helpful to collect and analyze property images taken from the street perspective. This may be especially useful in contexts such as ours where traditional hedonic data is thin or in applications where predictive power is the goal. Admittedly, estimating just the age of properties will not add much to many datasets outside the UK. Typically, the year of construction is part of both commercial and housing data already. However, automatic valuation models or hedonic models could improve by inserting architectural style, separate from age. In addition, other variables such as asset uniqueness, upkeep of the exterior, amount of greenery, presence of architectural features, some indicators of energy efficiency, type and intensity of usage and much more could be derived from images.

We are certain that more studies will tap into street-level images as an ubiquitous and powerful data source.

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 that 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, caused by 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, the hedonic estimates of architect classified styles are attenuated. 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 the 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 utilize building images taken from multiple angles to circumvent obstacles and to arrive at more reliable classifications. Street-level images are an 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

Table 1: Architectural Style Descriptions

Style	Era	Characteristics
Georgian	c1714–1837	sash windows, fan lights above doors, stucco on facades, wrought work grilles and railings
Early Victorian	c1837-c1870s	elaborate features such as carved barge boards or finials. sash windows more affordable and wider
Late Victorian/Edwardian	c1870s-1910	substantial bay windows, heterogeneous ornamentation, stained glass
Interwar	1918-1939	cost of building construction falls, Distinctive two-pane windows
Postwar	1950-1980	embrace of high-rise as well as low-rise housing. Facades vary between brick, tiling, pebbledash and render.
Contemporary	1980 - present	innovative and distinctive building techniques
Revival		contemporary buildings that emulate historic, mostly replica Victorian architecture

Notes: Prepared by members the Architecture Department at the University of Cambridge.

Figure 1: Examples of Architectural Styles

Georg.	Early Vic.	LateV/Ed	Interwar	Postwar	Contemp.	Revival

Notes: In these images, it is clear that much of the visual heterogeneity in style is due to differences in windows, doors, and rooflines. For example, there are clear differences across styles based on groupings and number of panes per windows and shapes of pediments, lintels, or fanlights above doors. Some characteristics stand out such as the presence of bay windows for Late V./Edwardian properties and angularity, asymmetry of windows in the Contemporary class, and roofing style and setback in the Postwar period.

20

Figure 2: Improving Camera Aim for Automatic Image Collections
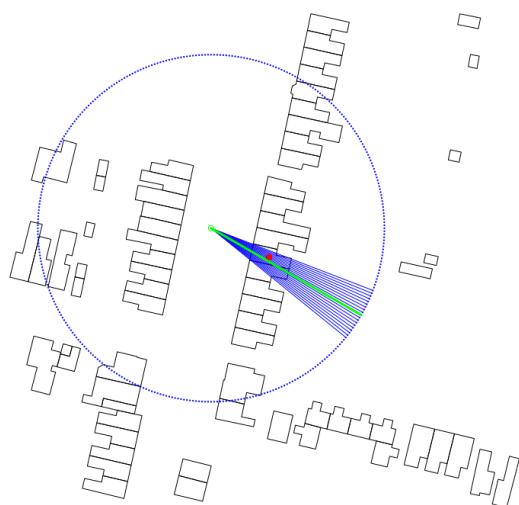
(a) Default Camera Aim

(b) Improved Aim

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 (a) the building of interest is not even visible in the picture. The image at (b) was taken with pan and zoom parameters, derived by our viewshed algorithm. Even for a terraced property, the structure contains minimal neighbor information, the roofline is captured, vehicles on the street are not visible the image is not obstructed by other structures. *Image source*: Google Street View.

21

Figure 3: 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 2: Confusion matrix for classification based on images of property and its nearest neighbor

Panel A: Model based on individual building and nearest neighbor

ML	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

ML	Georgian	Early Vic.	Late V./Edw.	Interwar	Postwar	Cont.	Revival
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

ML	Georgian	Early Vic.	Late V./Edw.	Interwar	Postwar	Cont.	Revival
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
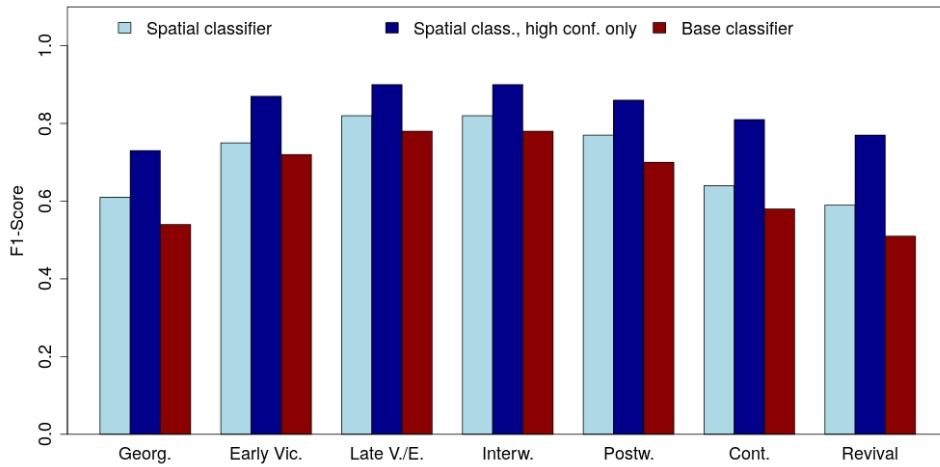
Notes: Cross-tabulation of out-of-sample predictions by the ML versus the architects' classification. *Recall* is the share of buildings from an architects' category being predicted correctly (diagonal in mid panel) and *Precision* is the share of buildings predicted to belong to a category that are indeed from that category. F_1-scores are the harmonic mean of Precision and Recall: F_1-*score = 2 Recall * Precision / (Recall + Precision).*

23

Table 3: ML classifications as dependent variables (Eq. 1 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 is lowest for the spatial ML predictor, especially when uncertain predictions are omitted (Column 4). This suggests that the spatial predictor performs best.

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 4: Prediction certainty: Herfindahl index from ensemble model

	Architects						
Mean Herf.	Georgian 0.88	Early Vic. 0.80	Late V. /Edw. 0.79	Interwar 0.77	Postwar 0.68	Cont. 0.70	Revival 0.70
ML	*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)	-0.25 (-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 5: 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 6: Differences in photo characteristics by correctly and incorrectly classified images

ML	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)	–
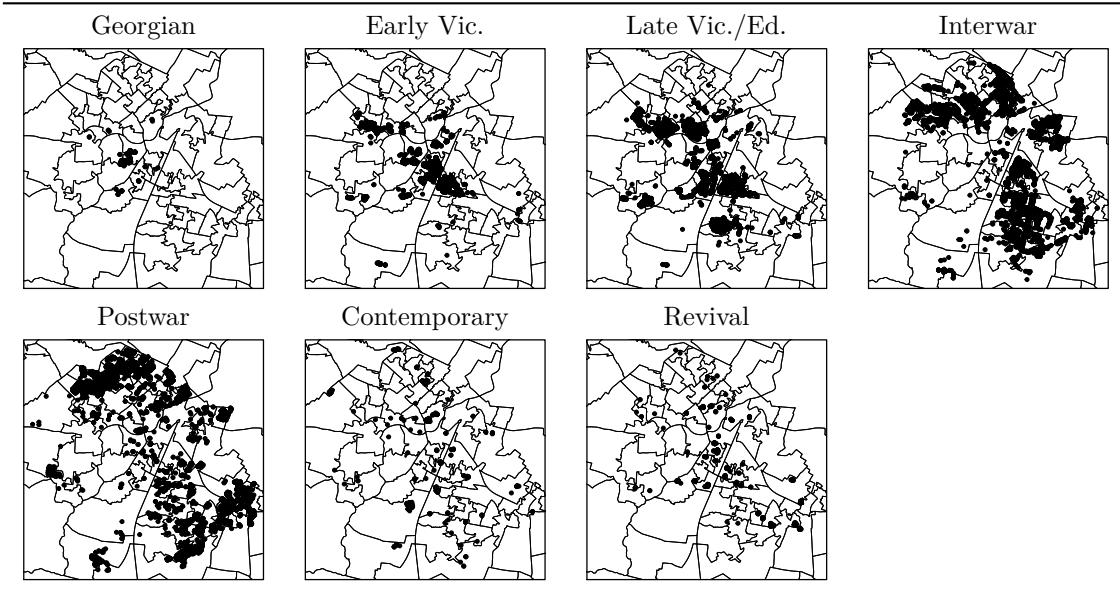
Notes: Cross-tabulation of photo characteristics by ML and architects' classifications. Off-diagonal elements ($ML_j - Architects_i$) describe the difference in measure intensity between correctly and incorrectly classified photos for a given classification by architects. For example, cell {Early Vic., 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.

26

Table 7: Summary statistics residential property transactions

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price	15,369	256,806.40	168,364.40	10,000	130,000	331,000	1,000,000
Year	15,369	2,004.97	6.60	1,995	1,999	2,010	2,018
Volume (m^3)	15,369	285.16	168.30	0.00	224.33	367.97	1,452.23
Area (m^2)	15,369	61.47	22.49	30.12	46.03	70.45	199.57
Dist. city center (m)	15,369	2,314.00	924.52	528.40	1,571.10	2,895.54	4,916.95
New	15,369	0.03	0.17	0	0	0	1
Type: detached	15,369	0.09	0.29	0	0	0	1
Type: semi-detached	15,369	0.36	0.48	0	0	1	1
Type: terraced	15,369	0.55	0.50	0	0	1	1
Georgian	15,369	0.02	0.13	0	0	0	1
Early Vic.	15,369	0.18	0.39	0	0	0	1
Late Vic./Edw.	15,369	0.25	0.43	0	0	0	1
Interwar	15,369	0.32	0.46	0	0	1	1
Postwar	15,369	0.17	0.38	0	0	0	1
Contemporary	15,369	0.04	0.18	0	0	0	1
Revival	15,369	0.03	0.17	0	0	0	1
Neigh: Georgian	15,369	0.02	0.13	0	0	0	1
Neigh: Early Vic.	15,369	0.17	0.38	0	0	0	1
Neigh: Late V./Edw.	15,369	0.26	0.44	0	0	1	1
Neigh: Interwar	15,369	0.32	0.47	0	0	1	1
Neigh: Postwar	15,369	0.18	0.39	0	0	0	1
Neigh: Contemporary	15,369	0.03	0.16	0	0	0	1
Neigh: Revival	15,369	0.02	0.14	0	0	0	1

Notes: Summary statistics for a sample of 15,511 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 ML classifications.

Figure 5: Spatial distribution of house sales, by architectural styles

Georgian	Early Vic.	Late Vic./Ed.	Interwar

Postwar	Contemporary	Revival

Notes: The locations of 15,369 housing transactions, styles classified by human experts, are marked by black dots. Neighborhood boundaries in the background.

28

Table 8: Hedonic Regression Estimates

	(1) base	(2) true	(3) pred.	(4) high conf.	(5) misclass.	(6) new	(7) neigh.	(8) style × neigh
	Dependent variable: *ln(price)*							
Constant	13.55***	13.18***	13.26***	13.01***	12.93***	14.52***	12.94***	12.97***
	(0.20)	(0.20)	(0.20)	(0.22)	(0.44)	(1.67)	(0.23)	(0.23)
ln(dist. city center)	−0.52***	−0.44***	−0.48***	−0.44***	−0.41***	−0.62***	−0.43***	−0.43***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.06)	(0.04)	(0.03)	(0.03)
Type: semi-detached	−0.11***	−0.12***	−0.12***	−0.10***	−0.14***	0.03	−0.10***	−0.10***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)	(0.01)
Type: terraced	−0.16***	−0.19***	−0.18***	−0.17***	−0.16***	−0.004	−0.17***	−0.17***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.08)	(0.01)	(0.01)
ln(area)	0.41***	0.39***	0.40***	0.41***	0.41***	0.34***	0.41***	0.41***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
ln(volume)	0.01***	0.01***	0.01***	0.01***	−0.004	0.01	0.01***	0.01***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)	(0.11)	(0.002)	(0.002)
New	0.19***	0.10***	0.18***	0.10***	0.13***		0.09***	0.08***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)		(0.03)	(0.03)
Base: Contemporary								
Georgian		−0.04	0.07***	−0.08**	−0.01		−0.08**	−0.27***
		(0.03)	(0.02)	(0.03)	(0.05)		(0.04)	(0.05)
Early Vic.		−0.13***	−0.01	−0.15***	−0.03		−0.11***	−0.11*
		(0.02)	(0.02)	(0.02)	(0.03)		(0.02)	(0.06)
Late Vic./Edw.		0.01	0.12***	0.004	0.03		0.003	0.12
		(0.02)	(0.02)	(0.02)	(0.03)		(0.02)	(0.09)
Interwar		−0.14***	−0.03**	−0.18***	−0.04		−0.15***	−0.18***
		(0.02)	(0.02)	(0.02)	(0.03)		(0.02)	(0.07)
Postwar		−0.22***	−0.09***	−0.26***	−0.11***		−0.19***	−0.23***
		(0.02)	(0.02)	(0.02)	(0.03)		(0.02)	(0.06)
Revival		0.04**	0.07***	0.03	0.08**	−0.06	0.02	0.07
		(0.02)	(0.02)	(0.02)	(0.03)	(0.08)	(0.03)	(0.06)
Base: Neigh. Contemporary								
Neigh: Georgian							−0.001	−0.43***
							(0.05)	(0.07)
Neigh: Early Vic.							−0.07**	−0.12***
							(0.03)	(0.04)
Neigh: Late V./Edw.							−0.004	−0.12
							(0.03)	(0.10)
Neigh: Interwar							−0.06**	0.01
							(0.03)	(0.05)
Neigh: Postwar							−0.10***	−0.10**
							(0.03)	(0.05)
Neigh: Revival							0.01	0.05
							(0.03)	(0.04)
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	15,369	15,369	15,369	11,627	3,063	317	11,627	11,627
Adjusted R^2	0.87	0.88	0.88	0.89	0.87	0.93	0.89	0.89
RMSE	0.235	0.225	0.228	0.219	0.235	0.119	0.218	0.217

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors are robust (White's estimator). Models (2) and (6) are estimated using the architects' classifications only; (3) is based on ML observations that have been automatically classified, while (4), (7) and (8) are based on high confidence ML predictions only; (5) uses architect's classifications on the subset of data where the architect and ML model style predictions match; (6) is estimated on sales of newly completed buildings. The interaction terms for (8) are presented in Table 9.

Table 9: Coefficients interaction terms: Building and neighborhood style

Neigh.	Building						
	Georg.	Early Vic.	Late V./Edw.	Interw.	Postw.	Cont.	Revival
Georg.	0.61**	0.41	0.77***	N/A	N/A	–	N/A
	(0.27)	(0.25)	(0.28)				
Early Vic.	0.27*	0.05	-0.11	0.03	0.10	–	0.07
	(0.16)	(0.11)	(0.13)	(0.10)	(0.11)		(0.12)
Late V./Edw.	0.36**	0.09	-0.01	0.15	0.06	–	0.10
	(0.17)	(0.12)	(0.14)	(0.11)	(0.12)		(0.14)
Interw.	N/A	-0.13	-0.17	-0.03	-0.03	–	-0.10
		(0.11)	(0.13)	(0.09)	(0.09)		(0.11)
Postw.	N/A	0.19	-0.19	0.03	0.05	–	-0.11
		(0.16)	(0.14)	(0.09)	(0.09)		(0.12)
Cont.	–	–	–	–	–	–	–
Revival	N/A	N/A	-0.45***	0.10	-0.06	–	-0.10
			(0.16)	(0.10)	(0.11)		(0.11)

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 8, Column 8 presents all other coefficients for this model.

Table 10: 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 frequently detected style on the same street, within 100m.

Table 11: Combined effect: Sum of direct, neighborhood and interaction coefficients

Neigh.	Building						
	Georg.	Early Vic.	Late V./Edw.	Interw.	Postw.	Cont.	Revival
Georgian	-0.09	-0.14	0.46	N/A	N/A	-0.43	N/A
Early Vic.	-0.11	-0.18	-0.11	-0.26	-0.25	-0.12	0.03
Late Vic./Edw.	-0.02	-0.14	0.00	-0.14	-0.29	-0.12	0.06
Interwar	N/A	-0.24	-0.04	-0.20	-0.25	0.01	-0.02
Postwar	N/A	-0.02	-0.17	-0.25	-0.29	-0.10	-0.14
Contemporary	-0.27	-0.11	0.12	-0.18	-0.23	0.00	0.07
Revival	N/A	N/A	-0.27	-0.03	-0.24	0.05	0.02

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.

Table 12: Hedonic Regression Estimates - Year x Neighborhood Fixed effects

	(1) base	(2) true	(3) pred.	(4) high conf.	(5) misclass.	(6) neigh.	(7) style × neigh
				Dependent variable: ln(price)			
Constant	13.55***	13.21***	13.25***	13.01***	13.30***	12.90***	12.89***
	(0.24)	(0.24)	(0.23)	(0.26)	(0.64)	(0.27)	(0.27)
ln(dist. city center)	−0.50***	−0.44***	−0.47***	−0.43***	−0.47***	−0.41***	−0.41***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.08)	(0.03)	(0.03)
Type: semi-detached	−0.12***	−0.12***	−0.12***	−0.10***	−0.14***	−0.10***	−0.10***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Type: terraced	−0.16***	−0.19***	−0.19***	−0.16***	−0.16***	−0.16***	−0.17***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
ln(area)	0.41***	0.39***	0.40***	0.41***	0.42***	0.41***	0.41***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)
ln(volume)	0.01***	0.01***	0.01***	0.01***	−0.01	0.01***	0.01***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)
New	0.15***	0.07***	0.14***	0.06*	0.17***	0.05	0.03
	(0.02)	(0.02)	(0.02)	(0.03)	(0.05)	(0.03)	(0.03)
Base: Contemporary							
Georgian		−0.02	0.08***	−0.08**	−0.07	−0.09**	−0.25***
		(0.03)	(0.03)	(0.03)	(0.07)	(0.04)	(0.07)
Early Vic.		−0.12***	0.003	−0.14***	−0.06	−0.11***	−0.24***
		(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.07)
Late Vic./Edw.		0.03	0.14***	0.02	0.002	0.01	0.17*
		(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.10)
Interwar		−0.13***	−0.01	−0.17***	−0.05	−0.14***	−0.21***
		(0.02)	(0.02)	(0.02)	(0.04)	(0.02)	(0.07)
Postwar		−0.21***	−0.07***	−0.25***	−0.14***	−0.20***	−0.23***
		(0.02)	(0.02)	(0.02)	(0.04)	(0.02)	(0.07)
Revival		0.06***	0.10***	0.05*	0.07	0.03	0.18***
		(0.02)	(0.02)	(0.03)	(0.05)	(0.03)	(0.07)
Base: Neigh. Contemporary							
Neigh: Georgian						0.01	−0.68***
						(0.05)	(0.13)
Neigh: Early Vic.						−0.06**	−0.11**
						(0.03)	(0.05)
Neigh: Late V./Edw.						0.002	−0.06
						(0.03)	(0.11)
Neigh: Interwar						−0.05*	0.02
						(0.03)	(0.05)
Neigh: Postwar						−0.09***	−0.12**
						(0.03)	(0.05)
Neigh: Revival						0.03	0.12**
						(0.03)	(0.05)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Neigh. dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year x Neigh dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Style Interaction terms	No	No	No	No	No	No	Yes
Observations	15,369	15,369	15,369	11,627	3,063	11,627	11,627
Adjusted R^2	0.87	0.88	0.88	0.89	0.88	0.89	0.89
RMSE	0.224	0.214	0.217	0.203	0.184	0.202	0.202

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors are robust (White's estimator). Model (2) and (6) are estimated using the architects' classifications; (3) is based on ML observations that have been automatically classified, while (4), (6) and (7) are based on high confidence ML predictions only; (5) uses architect's classifications on the subset of data where the architect and ML model style predictions match.

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 (2017). *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.

– (2020). "Beauty in the Eye of the Home-Owner: Aesthetic Zoning and Residential Property Values". In: *Real Estate Economics* 48.2, pp. 530–555. ISSN: 15406229. DOI: 10.1111/1540-6229.12204. URL: http://doi.wiley.com/10.1111/1540-6229.12204.

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.

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.