

# UC Santa Barbara

## UC Santa Barbara Previously Published Works

### Title

Beyond Spatial Auto-Regressive Models: Predicting Housing Prices with Satellite Imagery

### Permalink

<https://escholarship.org/uc/item/77c0438f>

### Authors

Bency, Archith J  
Rallapalli, Swati  
Ganti, Raghu K  
et al.

### Publication Date

2017

### DOI

10.1109/WACV.2017.42

Peer reviewed

# Beyond Spatial Auto-Regressive Models: Predicting Housing Prices with Satellite Imagery

Archith J. Bency<sup>1</sup>, Swati Rallapalli<sup>2</sup>, Raghu K. Ganti<sup>2</sup>, Mudhakar Srivatsa<sup>2</sup>, and B. S. Manjunath<sup>1</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA

<sup>2</sup>IBM T.J. Watson Research Center, Yorktown Heights, NY

## Abstract

*When modeling geo-spatial data, it is critical to capture spatial correlations for achieving high accuracy. Spatial Auto-Regression (SAR) is a common tool used to model such data, where the spatial contiguity matrix ( $W$ ) encodes the spatial correlations. However, the efficacy of SAR is limited by two factors. First, it depends on the choice of contiguity matrix, which is typically not learnt from data, but instead, is assumed to be known apriori. Second, it assumes that the observations can be explained by linear models.*

*In this paper, we propose a Convolutional Neural Network (CNN) framework to model geo-spatial data (specifically housing prices), to learn the spatial correlations automatically. We show that neighborhood information embedded in satellite imagery can be leveraged to achieve the desired spatial smoothing. An additional upside of our framework is the relaxation of linear assumption on the data. Specific challenges we tackle while implementing our framework include, (i) how much of the neighborhood is relevant while estimating housing prices? (ii) what is the right approach to capture multiple resolutions of satellite imagery? and (iii) what other data-sources can help improve the estimation of spatial correlations? We demonstrate a marked improvement of 57% on top of the SAR baseline through the use of features from deep neural networks for the cities of London, Birmingham and Liverpool.*

## 1. Introduction

Housing prices are important economic indicators of wealth and financial well-being in an urban scenario. In addition to house-specific metrics such as number of rooms and floors, square footage, and age, the location of houses also have been shown to affect valuations [15, 17]. Neighbourhood effects include factors such as taxation policies, availability of transportation and general amenities. In ad-

dition to housing prices, other socially and economically relevant metrics such as crime-rates and pollution levels [34, 28] also demonstrate spatial clustering. Hence, models designed to represent such geo-spatial data need to capture the underlying spatial correlations. Traditionally, spatially dependent phenomena as those mentioned above are described using Spatial Auto-Regressive (SAR) models.

The SAR model combines neighbourhood relationships between samples, and observed variables in a linear formulation to estimate spatially varying variables. The neighbourhood relationships are encoded in the form of a spatial contiguity matrix, and are often hand-designed with the help of domain expertise. The choice of the spatial contiguity matrix can lead to a trial and error process and leaves open the question of how to arrive at an optimal selection.

In this paper, we present a mechanism to learn the neighborhood relationship patterns from the data. To do so, we find that incorporating features for house locations learnt from satellite images is very effective. In recent times, multiple commercial real-estate listings websites [33, 32, 25, 30] store and display housing prices superimposed on satellite imagery from mapping services [20, 8]. The existence of such data makes large quantities of satellite images available with associated house prices. Satellite images provide a ‘bird’s eye view’ of a location and the neighbourhood it is situated in. In addition to the top-down appearance of a house, they also provide contextual information about the immediate and larger area of surroundings. We train Deep Convolutional Neural Networks (DCNNs) to discriminate between images learnt at different spatial scales corresponding to more and less affluent locations in a given city. The features learnt in the process are combined with house specific attributes through an estimator to arrive at a price estimate. The main contributions of this paper are:

- We present a method where neighbourhood information for geo-spatial samples is learnt implicitly through satellite image features

- We examine the impact of using neighbourhood information at multiple geo-spatial scales on housing prices estimation

The remainder of this paper is organized as follows. We present the related work in Section 2. Our approach is described in Section 3 with a brief review of the SAR model, deep feature extraction, multi-modal fusion and price estimation. In the subsequent Section 4, we characterize experiments conducted for price estimation and describe the data-sets, metrics and results. We discuss the results and conclude the paper in Section 5.

## 2. Related Work

We overview the related works by broadly dividing them into the following categories.

**Housing Price Estimation:** is a classical problem in the field of spatial econometrics [4, 23, 18]. These methods utilize attributes such as house square-footage, number of rooms, number of floors, age of the house, garage space etc. Extracting such detailed information for large data-sets would be a tedious task. Moreover, the spatial dependence of samples on each other is modelled using the SAR model. The choice of spatial contiguity matrix ( $W$ ) used by the SAR model is *hand designed* (as opposed to learning from the data) either using Delaunay Triangulation, k-Nearest Neighbours computation or Quasi-local correlation functions [18, 5]. While the work in [7] learns both sample level and spatially smooth manifold features from housing price data, it incorporates only non-visual features and needs fine-grained data such as type of heating and type of air conditioning amongst others which might be difficult to obtain on a large-scale for urban areas.

**Applications of Street View Imagery:** There has been increased attention in the computer vision community on the problem of urban scene analysis. Features learnt using Deep Convolutional Neural Networks [16, 27] have been shown to be effective at representing complex contextual information by learning from large-scale data-sets. [13, 22, 2, 3] investigated the correlation between visual features extracted from the street view imagery of cities and the high-level human perceptions on safety, wealth, directions to ubiquitous city landmarks and housing prices. These works have focused on utilizing street view images, which provide rich visual information in the immediate vicinity of houses, but do not describe a larger neighbourhood which is the case for satellite images.

**Applications of Satellite Imagery:** Satellite images have been analysed in the context of road detection [21, 10], predicting poverty [11], geolocalization [31], object detection

[6] and tracking [19]. In this work, we utilize images from this modality for the problem of housing price prediction. We further study the impact of using satellite images from different zoom levels on the accuracy of our models.

## 3. Estimating Housing Prices

### 3.1. Background: SAR model

SAR models are the traditional methods used to describe geo-spatial data. Dependent variables (to be estimated) are modelled as a weighted sum of dependent variable values of geo-spatial neighbours and the sample’s observed variables. Mathematically, the SAR model is represented as:

$$y = \rho W y + X \beta + \epsilon \quad (1)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

$y$  denotes the dependent variable of size  $n \times 1$ ,  $X$  of size  $n \times k$  represents the observed variables,  $W$  is the  $n \times n$  row normalized spatial contiguity matrix,  $\rho$  is the coefficient of spatial dependence for  $y$ ,  $\beta$  of size  $k \times 1$  signifies the influence of observed variables and  $\epsilon$  is the error term modelled as a zero mean Gaussian distribution.

The parameters of the model,  $\rho$  and  $\beta$  are learnt through Maximum Likelihood estimation [18]. The choice of  $W$  defines how neighbouring samples influence each other.  $W$  is constructed as a sparse matrix where  $W_{i,j} = 1$  for samples  $i$  and  $j$  which are neighbours. One criterion for samples to be neighbours is when they are within a distance of  $r_W$  of each other [5]. Another method of neighbourhood definition has been designed through Delaunay Triangulation [18]. Two samples which share an edge of a constructed triangulation are considered to be neighbours. It is apparent that the choice of  $W$  is highly dependent on domain expertise.

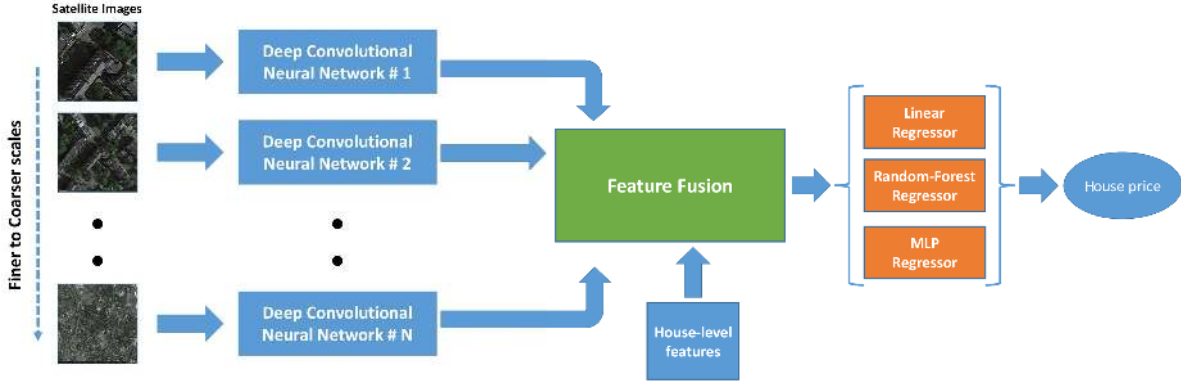
Under the condition of  $||\rho W|| < 1$ , equation 1 is rewritten as a power series expansion:

$$y = \sum_{i=0}^{\infty} \rho^i W^i (X \beta + \epsilon) \quad (2)$$

The equation can be interpreted as a decomposition of  $y$  in terms of increasing powers of  $W$ . Since  $W$  denotes spatial contiguity, terms with higher powers of  $W$  represent contribution of sample’s larger neighbourhoods in the value of  $y$ . In this work, we aim to emulate the effect of larger spatial neighbourhoods on dependent variables such as house prices through satellite images covering progressively larger geo-spatial areas, which provide a implicit and rich modality of information, instead of a hand-designed choice of  $W$ .

### 3.2. System Architecture

System architecture of the proposed method for estimating housing prices is presented in Figure 1. We detail the various components of the system below.



**Figure 1:** Satellite images of regions around test house samples are extracted from finer to coarser scales. Deep CNN features are extracted to get neighbourhood information and are fused with house-level explanatory variables through concatenation. The joint description of a house and its neighbourhood is used to estimate its price through regressors. *Best viewed in color.*

**Data Sources:** Our framework leverages multiple modalities of data. (i) *House Attributes:* First, we construct a database of house samples with latitude, longitude and house attributes from publicly available sources [33, 32]. House attributes are composed of number of bedrooms, bathrooms, reception rooms and floors. (ii) *Satellite Imagery:* Second, using the latitude and longitude coordinates, we query satellite images centered around the coordinate value [8, 20]. The images are sampled at different geo-spatial resolutions, keeping the image sizes at a constant value. The finer resolution scales result in images spanning the extent of individual houses, whereas coarser resolution images can span several city blocks or city districts.

**Feature Extraction From Satellite Imagery:** Our next goal is to be able to extract features from the satellite imagery to capture the neighborhood effects. To this end, we leverage Deep Convolutional Neural Networks (DCNNs). However, through our experiments, we found that training DCNNs directly for predicting housing prices is challenging. First, the models often converge slowly and overfit for the training data. This is likely due to the noise in the house prices and the aerial imagery not being able to distinguish between houses with slight differences in their asking prices. Therefore, to learn features that can generalize well to other data-sets, we use *transfer learning* and train the feature extraction pipeline for a similar but simpler problem. Through our experiments, we found that the binary classification problem of distinguishing between expensive and cheap houses, learns features that can generalize with excellent accuracy to other data-sets and other related tasks. The two classes we use for this task, are the top  $\delta\%$  and the bottom  $\delta\%$  of the training data-set in terms of price. Intuitively, expensive houses within a given city tend to exist in neighborhoods with larger backyard and green-space and water bodies such as ponds and swimming

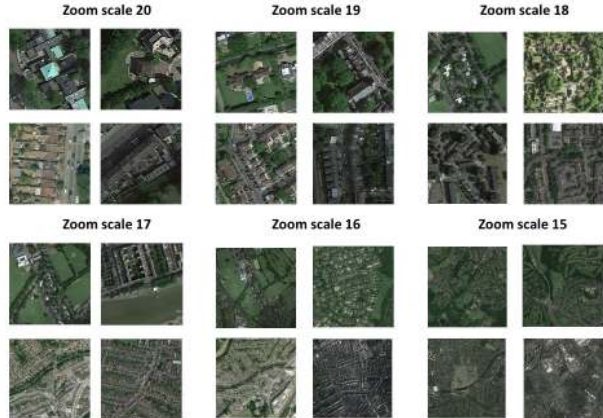
pools, whereas cheaper houses tend to be located in compact neighborhoods where the houses are adjacent to each other with concrete and roads occupying more space. These differences are apparent to the human eye in satellite images and we designed the choice of class definitions with this factor in mind. This choice of design for the classes is intended to enable networks to learn features which are sensitive to price variations. Some examples of satellite images used for training are provided in Figure 2.

Second, learning and combining features from different zoom levels of satellite imagery is non-trivial. One straightforward approach is to use a single network that can process all the zoom levels. However, our experiments showed that this mixes up the features, rendering poor accuracy. Therefore, as shown in Figure 1, we tackle this challenge by training a separate DCNN to learn features from each of the six zoom levels. Further details on how the DCNNs were trained are provided in Section 4.2.

**House Price Estimation:** The features extracted from the deep networks are concatenated with the house attributes into a feature vector  $x_{feat}$ . The resultant vector is then used to regress on housing prices through an estimator. In our experiments, we train multiple models, namely, (i) Linear, (ii) Random Forest and (iii) Multi-layer Perceptron regressors to understand and compare their effectiveness in estimating the housing prices.

### 3.3. Point of Interest Data

In order to validate the hypothesis that satellite images provide information regarding neighbourhoods for the task of price estimation, we also consider point of interest data. A point of interest is a location on a map which has economic, social or cultural value. Examples include fire-stations, restaurants, shopping centers, places of worship and bus stops. Each point of interest  $x$  is hence described



**Figure 2:** Examples of satellite images from the London city sale data-set. For each zoom value, the top row is constituted by examples from the top 10% of the data-set in terms of house price and the bottom row contains examples from the bottom 10%. Best viewed in color.

by latitude and longitude coordinates and a place type  $f_x$ . The set of all place types is denoted by  $P$ . In order to describe a house location  $h$  in terms of point of interest information, we generate a feature vector  $p_{h,r}$  which represents the number of instances of place type within a distance  $r$ . Mathematically,

$$\begin{aligned} p_{h,r} &= [p_{h,r,t}] \quad \forall t \in P \\ p_{h,r,t} &= \{ |x| \mid d_{h,x} < r, f_x = t \} \end{aligned} \quad (3)$$

where  $d_{h,x}$  is the haversine distance between locations  $h$  and  $x$  and  $|A|$  represents the cardinality of a set  $A$ . In our experiments, we utilize  $p_{h,r}$  as a feature vector describing the neighbourhood of the sample  $h$ . The parameter  $r$  serves as a hyper-parameter, for which the optimum value is deduced using cross-validation.

## 4. Experiments

### 4.1. Data-sets

**House price data-set:** We extract housing price listings from a web-based source. Each listing entry consists of house-level details such as house sale price or rental value in Pounds sterling (£), latitude, longitude, number of bathrooms, number of bedrooms, number of floors, number of reception rooms, listing status, street address, and a textual description. We extract the first eight attributes for each entry. The listing status for an entry can have two values, ‘rent’ or ‘sale’. Since the price ranges for the sale and rent subsets do not overlap, we conduct our experiments separately on each subset. For each of the data-sets, 10% of the samples are randomly picked without substitution and are designated as the *test* set and the rest are marked as the

City	Sale	Rent	Place of Interest
London	43037	55700	227800
Birmingham	3212	-	50306
Liverpool	5004	-	32878

**Table 1:** Number of sale and rent listings, and place of interest entries in city data-sets

*training* set.

We use data from the cities of London, Birmingham and Liverpool. The details regarding the number of samples in data-sets are presented in Table 1. We filter out samples with the top and bottom 2 % of the data-sets in terms of house sale price or rental value to prune out spurious or outlier entries. Due to relatively small number of samples in the ‘rent’ subset for Birmingham and Liverpool, we do not analyse rental price estimations for these cities. Distribution of prices in the London city data-set are presented in Figure 3, with price statistics for all data-sets shown in Table 2.

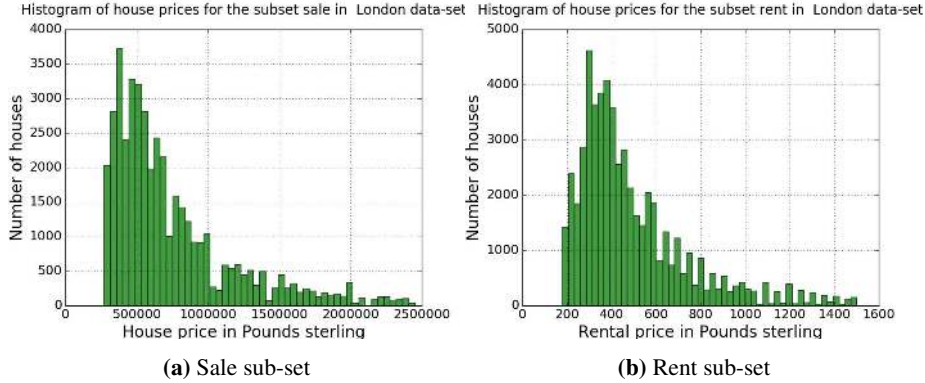
**Satellite Images:** Using the house coordinates, we extract RGB satellite images through the Google Maps [8] web service. Google Maps allows queries in the format of (latitude, longitude, zoom value, image size). Latitude and Longitude specify the center location of the satellite image, zoom values are integer values which specify the geo-spatial resolution of retrieved images. In this experiment, we specify queries with zoom values of 15, 16, 17, 18, 19 and 20, and the image size is fixed at 600×600. The images cover areas of 3.175, 0.794, 0.199, 0.048, 0.012 and 0.003 sq. kms. respectively. Considering the much larger average radius of the Earth, which is approximately 6371.0 km, we ignore effects of Earth’s curvature while calculating extents.

**Point of Interest Data:** In addition to satellite images, we have extracted place of interest data from the Google Places [9] web service. The Places service provides an interface to query places and businesses of interest described by tags within a specified radius from a latitude-longitude pair. Tags act as classes of places that are desired, with the service offering 86 pre-set tags.<sup>1</sup> Examples of tags include *cafe*, *beauty\_salon*, *clothing\_store* and *post\_office*. For each tag, our queries cover the entire area of the cities in our data-set. The total number of places of interest retrieved for each city is listed in Table 1. We construct a ball tree for each tag to facilitate efficient nearest-neighbour radial searches [14].

### 4.2. Deep Convolutional Neural Network training

In order to engineer features from the satellite images data-set, we train DCNNs to classify samples into the top and bottom  $\delta\%$  in terms of sale price in case of the ‘sale’

<sup>1</sup>The complete list of tags supported by the Google Places service is available at [https://developers.google.com/places/supported\\_types](https://developers.google.com/places/supported_types).



**Figure 3:** Distribution of house price and rental values for the city of London in our collected data-set.

City	Sale		Rent	
	Mean price (£)	Median price (£)	Mean price (£)	Median price (£)
London	743217.10	599997.04	492.02	415.00
Birmingham	181958.31	154999.94	-	-
Liverpool	151829.73	133252.35	-	-

**Table 2:** Statistics of sale and rent prices in city data-sets.

subset and rental price in case of the ‘rent’ subset as explained in Section 3. The value of  $\delta$  is set to 10 in our experiments to strike a balance between requirements of keeping reasonable number of samples available for training the DCNNs, and of keeping the image classes visually distinguishable. The choice of classes is also designed such that the features are sensitive to visual cues which distinguish between expensive and cheap houses and properties. Due to the limited size of our housing price data-sets, we fine-tune the Inception v3 [29] DCNN, which is a state-of-the-art image classifier trained on the larger Imagenet dataset [26]. Since both data-sets consist of natural images, the generic convolutional features learnt in the early layers are re-used and the final layers which are more task-specific are re-learned for the new task [12].

The final blocks of convolutional filters and fully connected layers (*mixed\_8x8x2048b*, *logits*) are re-trained, with rest of layers kept fixed to the values learnt from Imagenet. The fully connected layer *logits* is modified to generate features of dimension 256 which act as input to the final logistic classifier. In order to inhibit over-fitting on training data, we utilize  $L^2$  weight regularization, auxiliary loss layers, dropout and data-augmentation. The networks are trained for 50 epochs. We have utilized network fine-tune implementation for Inception v3 from the Tensorflow [1] toolbox.

Considering the number of samples in our city data-sets from Table 1, we train DCNNs for the London data-set alone and conduct experiments on the efficacy of applying

them on Liverpool and Birmingham in Section 4.5. We train six neural networks, one for each zoom value and keep a fixed learning rate of 0.001. The *train* subset is split in a 90:10 ratio into the classification *train* and *test* subsets. Table 3 shows that the classification accuracy of these networks, is above 90% for houses on sale and between 83% and 89% for houses put up for rent. Note that in this case, the chance performance is 50%. *The classification accuracy indicates that the networks are able to learn features that can distinguish between expensive and cheaper houses based on satellite images at each zoom level.*

For feature-extraction, we remove the logistic classifier layer from the neural network and the 256 dimensional features are used for price estimation. DCNN models were trained using Tensorflow [1] on a server configured with a Xeon E5-2630 CPU and a single NVIDIA Titan X GPU.

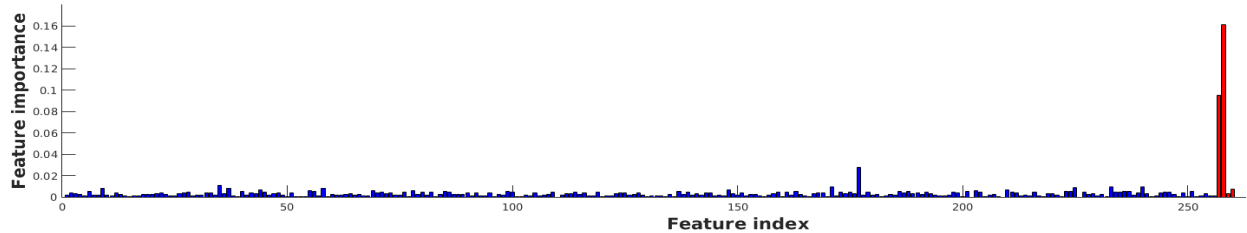
### 4.3. Estimators

We use three different types of estimators on features to regress on housing prices. The estimators used are (i) Linear, (ii) Random Forest (RF) and (iii) Multi-layer Perceptron (MLP) regressors. For RF regressors, we use 40 decision tree estimators. In the case of MLP regressors, we use 2 hidden layers with the number of nodes set as (500,100) for the London data-set and (50,10) for Liverpool and Birmingham data-sets. These choices of hyper parameters were arrived at by minimizing root mean square error on a random 90:10 train and validation split on the *train* subset. In order to account for stochasticity in training of RF and MLP regressors, we train 10 instances of the estimators and present mean and standard deviation for the result metrics in Section 4.5. All estimators were trained and tested using the Scikit-learn [24] python library.

### 4.4. Metrics

We use two standard regression metrics to report the efficacy of our proposed method in estimating house sale or





**Figure 4:** Mean Random Forest feature importance values as learnt for the London city sale dataset. The bars in blue represent Deep Features (DF) of the zoom value 20 and the bars in red represent Housing Attributes (HA). *Best viewed in color.*

Zoom value	15	16	17	18	19	20
Classification accuracy : Rent subset	88.96%	87.34%	86.04%	85.52%	84.32%	83.59%
Classification accuracy : Sale subset	90.36%	90.30%	90.49%	90.10%	90.30%	90.69%

**Table 3:** Performance for classifying between the top and bottom 10 % of samples in terms of house prices for different zoom values in the London data-set.

rental values: (i) Root Mean Square Error (RMSE) and (ii) Coefficient of Determination, also known as  $R^2$ .  $R^2$  measures how well the variation in data is explained by the regression model. If we consider  $y$  to be a vector of true price values and  $Y$  to be the estimated values,

$$R^2 = 1 - \frac{E[(Y - y)^2]}{E[(y - \bar{y})^2]} \quad (4)$$

with  $\bar{y}$  representing the mean value of  $y$ .  $R^2$  can take on values in the range  $(-\infty, 1.0]$ , with 1.0 indicating that the model is able to perform a perfect fit on test data.

#### 4.5. Results

**Comparison with SAR and other base-lines:** We compare performance of the proposed method with the SAR technique in Table 6. The observed variables matrix for SAR,  $X$  from equation 1, consists of the house attributes and point of interest features with  $r$  set to 2.0 km. We utilize an implementation of SAR from [18]. The choice of  $W$  is derived through Delaunay Triangulation (DT) and K-Nearest Neighbour (KNN). For the proposed method, concatenation of house attributes and deep features from satellite images from all zoom values is used to arrive at the results through Linear, Random Forest and MLP estimators. Even though the proposed method with Linear Regression and SAR are both linear models, we arrive at superior results through the usage of neighbourhood information in the form of satellite image features. Random Forest and MLP regressors further improve upon the estimation performance. A visualization of the comparison between predictions from the SAR model and the proposed method is presented in Figure 5.

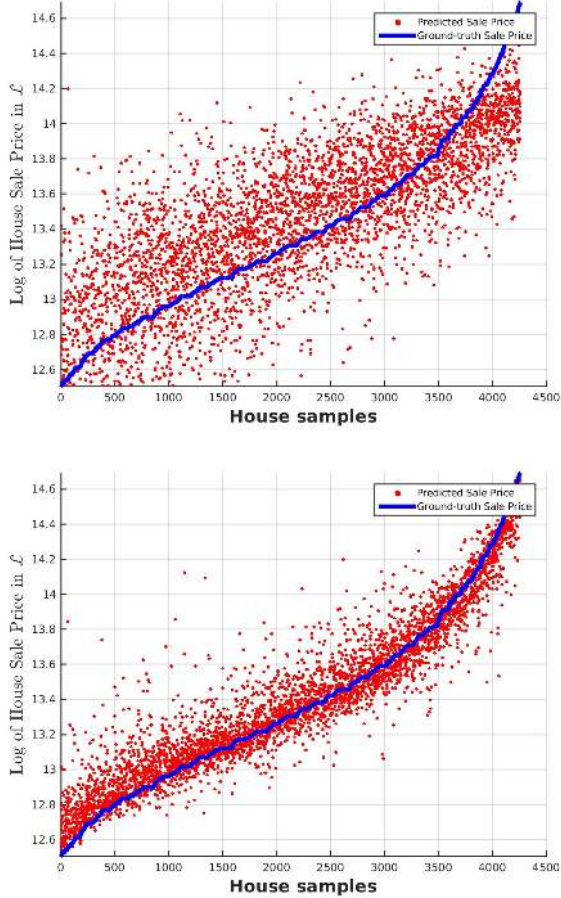
In Table 4, we compare our approach with [3], a recent method which estimates housing prices through house attributes (HA) and features extracted from street-level images through DCNNs. We use the data-set released by authors of the above work. Real estate data for houses in

Method	RMSE (\$)
Bessinger & Jacobs [3] : HA	<b>29365.00</b>
Proposed Method : HA	32108.91 $\pm$ 14.80
Bessinger & Jacobs [3] : IF	53727.00
Proposed Method : IF	<b>35188.72 <math>\pm</math> 7.98</b>
Bessinger & Jacobs [3] : HA + IF	28281
Proposed Method : HA + IF	<b>24439.64 <math>\pm</math> 11.63</b>

**Table 4:** Comparison of results for Housing price regression on the Fayette County house price data-set from [3]. Superior performance is indicated by lower RMSE values. HA and IF stand for House Attributes and Image Features respectively. IF for [3] are extracted from street view imagery whereas the proposed method utilizes satellite imagery.

Fayette County, Kentucky, USA are provided with details regarding house location and attributes such as *Tax rates*, *Acre*s and *Total Rooms* along with a train-test split. As the networks used for feature extraction in [3] are fine-tuned versions of the VGG-16 network [27], we also utilize the same architecture with training procedure as described in Section 4.2. We demonstrate superior performance through a 13.5% reduction in RMSE through using HA + IF features. In the case of using Image Features, which in our case are the features extracted through DCNNs on the satellite images extracted at different zoom values, we achieve a larger reduction of 34.5% in RMSE. The results show that features through multi-scale satellite imagery are able to better explain housing price variations. Random Forest regressors with 720 estimator trees were used to arrive at this result.

**Effect of Feature Combinations:** We also experiment with using different feature combinations at the feature fusion stage of our approach. We leverage three classes of features: (i) deep features (DF) extracted from satellite images, (ii) house attributes (HA), and (iii) place of interest (POI) features. The results in terms of RMSE and  $R^2$  are listed in Table 7. HA features are unable to capture price



**Figure 5:** Top: Logarithm of SAR model predicted sale housing prices for the London city data-set is shown as a scatter plot. Bottom: Logarithm of proposed method predicted sale housing prices for the same data-set with House Attribute and Deep Features using Random Forest regression.

variations, as they characterise the house itself, but not its neighborhood. Estimators using DF or POI features, which capture the neighborhood information, are able to improve the prediction. When we combine HA with DF or POI features for estimation, a large improvement over the previous configurations is seen. *This observation indicates that house level attributes and neighborhood level features such as DF and POI are complementary in capturing price variations.* An additional observation is that combining DF and POI features does not lead to any significant advance in price estimation over the case where either of them are utilized, demonstrating that *DF and POI features are highly correlated* in the context of price estimation. The results also highlight the difference between ‘sale’ and ‘rent’ sub-markets. For rental properties, HA by itself is able to explain the variation of prices, in contrast to the houses put up

Deep Feature Zoom Values	London	
	Rent	Sale
RMSE (£)		
HA + DF (20)	81.20 ± 0.71	156821.14 ± 1534.30
HA + DF (19, 20)	80.65 ± 0.47	157200.04 ± 1617.62
HA + DF (18, 19, 20)	79.14 ± 0.63	155761.95 ± 1512.96
HA + DF (17, 18, 19, 20)	<b>79.34 ± 0.54</b>	156010.72 ± 1209.60
HA + DF (16, 17, 18, 19, 20)	79.59 ± 0.24	144524.24 ± 1050.83
HA + DF (15, 16, 7, 18, 19, 20)	79.78 ± 0.63	<b>127303.06 ± 1634.04</b>
$R^2$		
HA + DF (20)	0.8905 ± 0.0019	0.8647 ± 0.0026
HA + DF (19, 20)	0.8925 ± 0.0012	0.8640 ± 0.0028
HA + DF (18, 19, 20)	0.8960 ± 0.0016	0.8665 ± 0.0026
HA + DF (17, 18, 19, 20)	<b>0.8955 ± 0.0014</b>	0.8661 ± 0.0021
HA + DF (16, 17, 18, 19, 20)	0.8948 ± 0.0006	0.8851 ± 0.0016
HA + DF (15, 16, 7, 18, 19, 20)	0.8946 ± 0.0008	<b>0.9108 ± 0.0022</b>

**Table 5:** RMSE and  $R^2$  for different neighbourhood zoom values with Random Forest regression. Lower values for RMSE and higher values for  $R^2$  indicated superior performance with the method with best results highlighted in bold. HA and DF stand for House Attributes and Deep Features respectively.

for sale. *This indicates that transactions carried out for relatively short-term usage of houses place more value on the amenities of the house itself rather than its neighborhood.* We have also visualized the relative frequency of feature selection with Random Forest estimators in Figure 4. HA features are the most selected from amongst the features, but as can be seen in Table 7, adding DF features leads to improvement in estimator performance.

**Effect of neighbourhood size:** We next conduct experiments on the effects of including information from progressively larger neighbourhoods of house samples through satellite images. The zoom values for Google Map queries are integer values ranging from 15 to 20. We present RMSE and  $R^2$  results using the Random Forest estimator for the London data-set in Table 5. *The results show that the proposed method is able to better predict price by including larger neighbourhood contexts.* This is a significant observation, because most often, the spatial contiguity matrix specified in the SAR model only captures local neighborhood information, which is not sufficient to capture all the spatial relationships.

**Deep feature extraction across cities:** As described in Section 4.2, the DCNNs were trained using data from London data-set due to limited number of samples in Birmingham and Liverpool data-sets. From Table 6, we can observe that the features extracted through networks trained on a specific data-set are effective in estimating housing prices in a different city. *This implies that satellite image feature extraction is generic across cities in the same broad geographical region, the British Isles.* Further studies could examine how effective such feature extraction schemes would be for cities in more contrasting regions.



Method	London		Birmingham	Liverpool
	Rent price	Sale price	Sale price	Sale price
	RMSE (£)			
SAR (DT)	159.18	282989.82	52739.86	58075.00
SAR (K-NN, K = 10)	159.85	284324.75	58494.00	53249.85
Linear	143.75	259411.88	38316.55	39395.46
Random Forest	80.04 ± 0.51	127328.66 ± 1204.43	27868.22 ± 968.52	31412.19 ± 565.78
MLP	<b>74.52 ± 3.10</b>	<b>116639.78 ± 2941.28</b>	<b>20837.38 ± 140.60</b>	<b>29690.90 ± 408.67</b>
	$R^2$			
SAR (DT)	0.5794	0.5596	0.5884	0.5231
SAR (K-NN, K=10)	0.5759	0.5554	0.5804	0.5957
Linear	0.6569	0.6298	0.7824	0.8165
Random Forest	0.8936 ± 0.0014	0.9108 ± 0.0017	0.8848 ± 0.0081	0.8833 ± 0.0042
MLP	<b>0.9077 ± 0.0078</b>	<b>0.9251 ± 0.0038</b>	<b>0.9356 ± 0.0009</b>	<b>0.8958 ± 0.0029</b>

**Table 6:** RMSE and  $R^2$  for SAR and proposed housing price prediction models. Lower values for RMSE and Higher values for  $R^2$  indicated superior performance with the method with best results highlighted in bold.

Method	London			
	Rent price RMSE (£)	Sale price RMSE (£)	Rent price $R^2$	Sale price $R^2$
	<b>Random Forest Regression</b>			
HA	130.10 ± 0.49	366954.09 ± 130.14	0.7190 ± 0.0020	0.2594 ± 0.0005
POI	200.79 ± 0.06	202220.03 ± 1551.227	0.3307 ± 0.0004	0.7751 ± 0.0034
DF	136.59 ± 0.63	213617.63 ± 1472.42	0.6903 ± 0.0029	0.7490 ± 0.0035
HA + POI	<b>72.24 ± 0.69</b>	<b>108962.85 ± 769.53</b>	<b>0.9133 ± 0.0017</b>	<b>0.9347 ± 0.0009</b>
HA + DF	80.04 ± 0.51	127328.66 ± 1204.43	0.8936 ± 0.0014	0.9108 ± 0.0017
DF + POI	135.72 ± 0.51	211056.00 ± 849.26	0.6942 ± 0.0023	0.7550 ± 0.0020
	<b>Multi-layer Perceptron Regression</b>			
HA	204.48 ± 0.49	373608.36 ± 67.50	0.3059 ± 0.0033	0.2323 ± 0.0003
POI	183.93 ± 3.69	335122.36 ± 4552.90	0.4382 ± 0.0223	0.3822 ± 0.0168
DF	129.38 ± 2.74	190841.13 ± 1343.46	0.7220 ± 0.0119	0.7997 ± 0.0028
HA + POI	116.91 ± 3.68	213356.72 ± 5795.78	0.7729 ± 0.0142	0.7495 ± 0.0137
HA + DF	<b>74.52 ± 3.10</b>	<b>116639.78 ± 2941.28</b>	<b>0.9077 ± 0.0078</b>	<b>0.9251 ± 0.0038</b>
DF + POI	128.07 ± 1.63	190945.77 ± 1443.09	0.7277 ± 0.0069	0.7995 ± 0.0030

**Table 7:** RMSE and  $R^2$  for different feature combinations with Random Forest and Multi-layer perceptron regression. Lower values for RMSE and higher values for  $R^2$  indicated superior performance with the method with best results highlighted in bold. HA, POI and DF stand for House Attributes, Point of Interest and Deep Features respectively.

## 5. Discussion and Conclusion

From the experiments, we can see that joint representations of houses in terms of their individual attributes (HA) and the neighbourhood around them improves the accuracy of housing price prediction. This result is reinforced by conclusions from earlier works such as [7]. PoI and DF features have been consistently shown to be positively correlated in their effects on house prices across estimators and city datasets. Hence, multi-scale DCNN-derived features from publicly available satellite imagery could be used in place of PoI data, which are usually of proprietary nature and require explicit annotations regarding local businesses.

As we can see from the results, utilizing deep features from satellite images at different geo-spatial resolutions leads to comparable or superior performance compared to using either latitude-longitude information explicitly in SAR models or utilizing Point of interest features. Also, employing information from images of larger areas

surrounding house samples leads to improved accuracy in housing price estimation.

## 6. Acknowledgements

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053 (the ARL Network Science CTA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

## References

- [1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al.

- Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
- [2] S. M. Arietta, A. A. Efros, R. Ramamoorthi, and M. Agrawala. City forensics: Using visual elements to predict non-visual city attributes. *IEEE transactions on visualization and computer graphics*, 20(12):2624–2633, 2014.
  - [3] Z. Bessinger and N. Jacobs. Quantifying curb appeal. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 4388–4392. IEEE, 2016.
  - [4] S. C. Bourassa, E. Cantoni, and M. Hoesli. Spatial dependence, housing submarkets, and house price prediction. *The Journal of Real Estate Finance and Economics*, 35(2):143–160, 2007.
  - [5] Y. Chen. On the four types of weight functions for spatial contiguity matrix. *Letters in Spatial and Resource Sciences*, 5(2):65–72, 2012.
  - [6] G. Cheng and J. Han. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117:11–28, 2016.
  - [7] S. Chopra, T. Thampy, J. Leahy, A. Caplin, and Y. LeCun. Discovering the hidden structure of house prices with a non-parametric latent manifold model. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 173–182. ACM, 2007.
  - [8] Google Inc. *Google Maps API*, 2016 (accessed October 6th, 2016). <https://developers.google.com/maps/>.
  - [9] Google Inc. *Google Places API*, 2016 (accessed October 6th, 2016). <https://developers.google.com/maps/>.
  - [10] J. Hu, A. Razdan, J. C. Femiani, M. Cui, and P. Wonka. Road network extraction and intersection detection from aerial images by tracking road footprints. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4144–4157, 2007.
  - [11] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353:790–794, 2016.
  - [12] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller. Recognizing image style. In *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014.
  - [13] A. Khosla, B. An, J. J. Lim, and A. Torralba. Looking beyond the visible scene. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Ohio, USA, June 2014.
  - [14] A. M. Kibriya and E. Frank. An empirical comparison of exact nearest neighbour algorithms. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 140–151. Springer, 2007.
  - [15] K. Kockelman. Effects of location elements on home purchase prices and rents in san francisco bay area. *Transportation Research Record: Journal of the Transportation Research Board*, (1606):40–50, 1997.
  - [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
  - [17] D. J. Krupka and D. Noonan. Neighborhood dynamics and the housing price effects of spatially targeted economic development policy. *SSRN Working Paper Series*, 2009.
  - [18] J. P. LeSage and R. K. Pace. *Introduction to Spatial Econometrics (Statistics, textbooks and monographs)*. CRC Press, 2009.
  - [19] L. Meng and J. P. Kerekes. Object tracking using high resolution satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1):146–152, 2012.
  - [20] Microsoft Corporation. *Bing Maps: Developer Resources*, 2016 (accessed October 6th, 2016). <https://www.microsoft.com/maps/developer-resources.aspx>.
  - [21] V. Mnih and G. E. Hinton. Learning to detect roads in high-resolution aerial images. In *European Conference on Computer Vision*, pages 210–223. Springer, 2010.
  - [22] V. Ordonez and T. L. Berg. Learning high-level judgments of urban perception. In *European Conference on Computer Vision*, pages 494–510. Springer, 2014.
  - [23] L. Osland. An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, 2010.
  - [24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
  - [25] Redfin. Redfin: Real estate, homes for sale, mls listings, agents, 2016 (accessed October 6th, 2016). <https://www.redfin.com/>.
  - [26] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
  - [27] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.
  - [28] W. Steenbeek and J. R. Hipp. A longitudinal test of social disorganization theory: Feedback effects among cohesion, social control, and disorder. *Criminology*, 49(3):833–871, 2011.
  - [29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. *arXiv preprint arXiv:1512.00567*, 2015.
  - [30] Trulia. Trulia: Real estate listings, homes for sale, housing data, 2016 (accessed October 6th, 2016). <https://www.trulia.com/>.
  - [31] S. Workman, R. Souvenir, and N. Jacobs. Wide-area image geolocalization with aerial reference imagery. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1–9, 2015.
  - [32] Zillow Group. Zillow: Real estate, apartments, mortgages & home values, 2016 (accessed October 6th, 2016). <http://www.zillow.com/>.
  - [33] Zoopla PLC. *Zoopla: Search Property to Buy, Rent, House Prices, Estate Agents*, 2016 (accessed October 6th, 2016). <http://www.zoopla.co.uk>.

- [34] L. M. Zwack, C. J. Paciorek, J. D. Spengler, and J. I. Levy. Modeling spatial patterns of traffic-related air pollutants in complex urban terrain. *Environmental Health Perspectives*, 119(6):852, 2011.