QUANTIFYING CURB APPEAL

Zachary Bessinger *Nathan Jacobs*

Department of Computer Science, University of Kentucky
{zach, jacobs}@cs.uky.edu

ABSTRACT

The curb appeal of a home, which refers to how attractive it is when viewed from the street, is an important decision-making factor for many home buyers. Existing models for automatically estimating the price of a home ignore this factor, instead focusing exclusively on objective attributes, such as number of bedrooms, the square footage, and the age. We propose to use street-level imagery of a home, in addition to the objective attributes, to estimate the price of the home, thereby quantifying curb appeal. Our method uses deep convolutional neural networks to extract informative image features. We introduce a large dataset to support an extensive evaluation of several approaches. We find that using images and objective attributes together results in more accurate home price estimates than using either in isolation. We also find that representations learned for scene classification tasks are more discriminative for home price estimation than those learned for other tasks.

***Index Terms*—** image understanding, computational aesthetics, neural networks, image feature representation

Fig. 1: Appearance provides a strong cue for estimating the value of a home. The same houses before (top) and after (bottom) renovations to increase their "curb appeal." Updated architecture, paint, and lawn care are essential for high curb appeal. (Images courtesy of Houzz.)

1. INTRODUCTION

There is significant interdisciplinary research related to aesthetics of urban spaces and their impact on human behavior. Perhaps one of the most well-known theories is Wilson's *broken windows theory* [1] that suggests if a home or building has broken windows, then it is highly probable that the neighborhood has high crime, vandalism, and disorder. Conversely improving lighting and adding sidewalks can reduce crime, improve health, and increase happiness [2, 3]. We approach the issue of aesthetics from an economic perspective, focusing on its impact on the monetary value of a property.

The attractiveness of a home's exterior is colloquially referred to as its *curb appeal*, a significant factor in a home purchasing decision and, consequently, the value of a home. A large industry, including contractors, television shows, magazines, and blogs, are devoted to ways to improve curb appeal. Some examples of changes that can improve curb appeal include lawn care, repainting, adding a pergola, or changing the mailbox. Figure 1 shows the appearance of two homes before and after exterior renovations. These renovations have arguably increased the curb appeal of the home, which would likely increase its sale price. This work explores learning-based methods for quantifying the curb appeal of the home, through estimating its price from street-level imagery.

Our approach differs significantly from previous work on automatic home price estimation. Existing approaches are based on a combination of objective attributes, such as the square footage, the number of bedrooms, and the price of homes with similar attributes in the area [4, 5]. This means that they are unable to account for recent changes in the curb appeal of the home, which a human appraiser would take into consideration. We propose to use the appearance of a home, captured by street-level imagery, as a factor in predicting the fair-cash value of a home. In experiments on a diverse urban area, we find that this results in an average of $\$2\,401$, or 6.14%, reduction in error.

The main contributions of this work are: 1) a large-scale evaluation of applying learned image features to the domain of understanding home aesthetics, 2) a joint model that incorporates images and metadata to improve home price estimation, and 3) an extensive evaluation demonstrating our joint model outperforms independent metadata and image models.

2. RELATED WORK

Our work introduces a novel method for using imagery to understand homes in an urban area and is a special case of image attribute estimation. The remainder of this section provides an overview of work in these two areas.

2.1. Estimating Image Attributes

Early work on estimating image attributes of scenes, which are visible semantically meaningful properties, include the works of Patterson et al. [6], Su et al. [7], and Parikh et al. [8]. These early works relied on BoW models and semantic segmentation to learn semantic attributes. More recently Zhou et al. [9, 10] have leveraged recent advances in convolutional neural networks to create benchmark datasets useful for improving the task of scene classification. Once a scene can be classified into a number of categories, that information can be used as prior knowledge towards higher-level attribute estimation. Laffont et al. [11] estimate a set of contextual attributes related to weather and seasons from outdoor images and apply appearance transfer to effectively synthesize weather effects on other outdoor images. Glasner et al. [12] learn a temperature attribute to transform a camera into a crude temperature sensor. Our work extends previous works of learning attributes for outdoor scenes by estimating the attribute of curb appeal.

2.2. Using Imagery to Understand Urban Areas

Recently there has been an increased interest in applying computer vision techniques to learn attributes for understanding urban areas. The most heavily researched attribute in this research subject is automatically assessing the safety of a region [13, 14, 15]. Porzi et al. [16] leverage the power and speed of convolutional neural networks (CNNs) to rank safety in an end-to-end manner from Google Street View images. Other recent works have learned how certain aspects of cities, such as their architecture [17] and architectural evolution [18], can be used as cues for city recognition. Salesses et al. [19] and Quercia et al. [20] have successfully estimated attributes of wealth and beauty of a city based on images obtained from Google Street View and social media. Our work takes a similar approach, but with a more specific focus; we quantify curb appeal and use it to improve our ability to estimate the price of a home.

3. DATASET

We constructed a dataset of publicly available residential housing data from the Fayette County Property Valuation Administrator (PVA) [21]. Each record in the dataset has 15 metadata attributes including fair-cash value, number of bedrooms/bathrooms, and year built. We use the fair-cash value, which is determined by a human appraiser, as the price of the home. Each home also has at least one front-facing 640×480 image captured by an appraiser. Figure 2 shows example images from our dataset. These images are captured at varying angles and weather conditions.

Fig. 2: Images from our dataset, showing the diversity of architectural styles, landscaping, and lighting conditions.

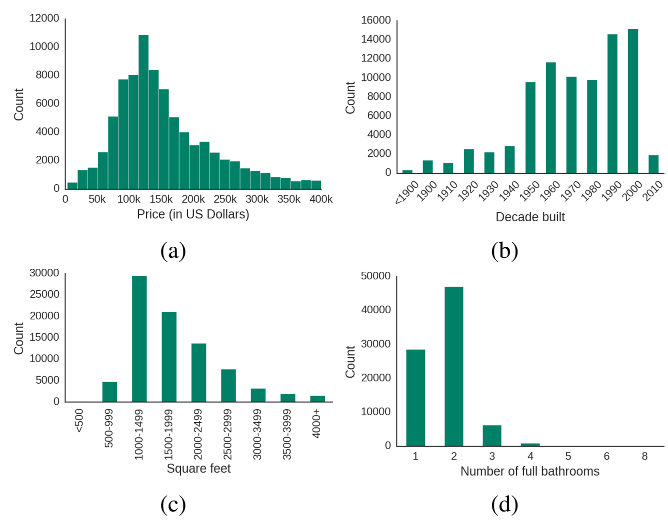

Fig. 3: The distribution of homes in our dataset with respect to four different attributes.

There are a small quantity of houses that are outliers in terms of price. We alleviate this by thresholding for price and keeping all homes valued at \$400 000 or less. After filtering, our final dataset contains 83 140 records. Histograms of several objective attributes are visualized in Figure 3. The home prices have mean $\mu = \$154\,829$ and standard deviation $\sigma = \$76\,082$.

4. METHODS

We propose a novel method that accounts for curb appeal by jointly using metadata and image features to predict the value of a home. We compare with two baseline methods that directly predict the price, one that uses only metadata features and one that uses only image features. We generate price prediction models using three regression techniques: linear regression, ridge regression, and random forest regression. The notation, $P(x)$, refers to a price prediction model, given some input data, x. The following subsections describe the features used to construct each price prediction model. Evaluation results are presented in Section 5.

4.1. Home Prices using Metadata

In our metadata-only model, $P(M)$, we ignore the available image data and directly predict the price of a home from its objective attributes. We search all possible subsets of attributes and fit linear regression models to predict price. We select the optimal model with dimensionality eight that maximizes the R^2 score because including additional attributes does not significantly reduce the error. This model uses the following attributes: the number of acres, bedrooms, bathrooms, year built, residential square footage, total fixtures, basement size, and garage size. These eight objective attributes are used as our metadata feature representation in all experiments. Each attribute is scaled to have mean $\mu = 0$ and variance $\sigma^2 = 1$ and a metadata-only model is learned for each of the three regressors.

4.2. Home Prices using Images

Our image-only model, $P(I)$, ignores the metadata attributes and predicts the price of the home using only an exterior photograph. We investigate a variety of feature representations extracted from CNNs, which have recently been shown [15, 22] to perform better than hand-crafted features. Since our images are outdoor, lighting factors such as the sun, shadows, and seasonal changes strongly affect the image appearance. Therefore, it is imperative that our visual features be invariant to lighting and weather conditions. Motivated by this, we perform a large-scale evaluation of existing CNN architectures to find a feature representation optimal for predicting price. Using each of the three regressors, we learn models using a variety of image features and select the features that minimize the root mean squared error. The evaluation for selecting the best image features is described in Section 5.2.

4.3. Home Prices using Metadata and Images

Our proposed model, $P(M) + C(I)$, combines metadata attributes and image features to predict the price of the home. This model estimates the home price using two components: the predicted home price and the curb appeal modifier. The metadata-only model, $P(M)$, defined in Section 4.1 is used to directly predict the home price. The image features used in the image-only model in Section 4.2 to directly predict the price now predict the curb appeal modifier, $C(I)$, which adjusts the predicted home price, positively or negatively, based on its curb appeal.

Table 1: RMSE (in U.S. Dollars) for each price prediction model. The best result for each column is bolded. Lower is better.

	Linear	Ridge	Random Forest
$P(M)$	\$37 058	\$37 127	\$29 365
$P(I)$	\$53 575	\$53 565	\$53 727
$P(M) + C(I)$	**\$34 538**	**\$34 606**	**\$28 281**

5. EVALUATION

We present a quantitative and qualitative evaluation of the proposed methods using the dataset described in Section 3. We uniformly at random select 80% of the homes as a training set, reserving the rest for testing. Qualitative results are presented for the top-performing model, showing homes whose image had significant impact, both positive and negative, on the predicted price.

5.1. Implementation Details

Our proposed method is implemented in Python using Scikit-Learn [23] for the regressors and Caffe [24] for image feature extraction. The optimal hyperparameters for each regressor is learned using hyperopt [25]. Optimization is done by applying 5-fold cross validation on the training set and minimizing the average root mean squared error. Our source code will be made available pending publication.

5.2. Image Feature Selection

To find optimal image features for predicting price, we learn several regression models trained on features extracted from ten convolutional networks. The networks are trained on three sets of training data, ImageNet [26], Places [9], and Places2 [10] using four different architectures, AlexNet [27], GoogLeNet [28], and VGG-16/VGG-19 [29]. We find that features extracted from similar network architectures trained on Places2 and ImageNet significantly differ. Comparing two random forests to predict price with similar hyperparameters using VGG-16 FC6 features, one trained on Places2 and the other trained on ImageNet, Places2 trained image features reduce the root mean squared error (RMSE) by \$6 140 more than ImageNet features. Similarly, using Places2 FC6 and FC8 features instead of Places FC6 and FC8 features reduce the RMSE by \$1 848 and \$1 636, respectively. Since the

(a) (b)

Fig. 4: Homes where the curb appeal modifier in our joint model has a strong negative (a) and strong positive (b) impact on estimated price, respectively.

difference in dimensionality between FC6 and FC8 is greater than a factor of 10 and FC8 is the layer that corresponds to semantic labels, we prefer to use the FC8 layer. To this end, we select the FC8 layer of the top performing network architecture, Places2 VGG-16, from which to extract image features. These features are used as our image feature representation for all further evaluations.

5.3. Quantitative Results

We report the RMSE for each price estimation model on the testing set in Table 1. We find that the random forest models achieve the highest performance compared to other tested regressors. Our joint model using random forests for predicting the price component with metadata and the curb appeal modifier with image features shows improvement over the baseline metadata-only and image-only models, achieving a RMSE of $\$28\,281$. This is a relative improvement of 4.0% over our metadata-only model and 90.0% improvement over our image-only model using similar regressors. For all regression methods, we observe that incorporating curb appeal into the price prediction yields more accurate results.

Finding the optimal hyperparameters for ridge regression involves optimizing over α. In the metadata-only model $\alpha_M = 44.259$ and the image-only model $\alpha_I = 16.791$. The joint model metadata component uses α_M and the curb appeal modifier uses $\alpha_C = 83.918$. For random forest regression we optimize over the number of estimators and the minimum number of features to create a split. $P(M)$ in uses 794 estimators and considers 4 features for each split. $P(I)$ uses 1188 estimators and considers 167 features for each split. In the joint model, $C(I)$ uses 522 estimators and 157 features for each split.

5.4. Influence of Curb Appeal on Price

To better understand how the curb appeal modifier in our joint model influences estimated price, we explore the relationship between semantic image labels corresponding to the Places2 VGG-16 FC8 layer and predicted home price. Figure 4 shows several homes whose prices are strongly affected by their curb appeal. The homes in Figure 4a were significantly devalued by incorporating curb appeal. The opposite effect happens in Figure 4b, where incorporating curb appeal significantly increased the estimated price. Elements of high curb appeal, green grass and kempt appearance, are clearly visible in Figure 4b and strongly influence the home's value.

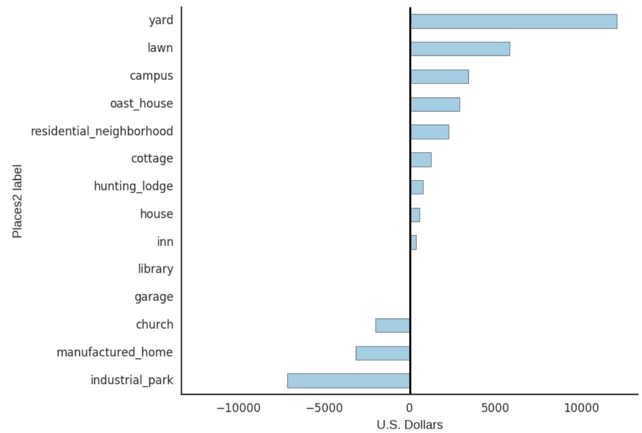

Fig. 5: Mean predicted curb appeal modifier for homes grouped by semantic label from Places2 FC8 layer.

To further understand semantics, we aggregate the semantic labels from the test set and filter them for predictions that occurred more than 100 times. For each label, we calculate the mean predicted curb appeal modifier and show it in Figure 5. Scenes predicted as "yard" and "lawn" add $\$12\,105$ and $\$5\,823$ to the price of the home, on average. Similarly, scenes predicted as "industrial park" and "manufactured home" subtract $\$7\,152$ and $\$3\,162$ from the price of the home, on average. This further supports the hypothesis that a kempt home with natural beauty is more valuable.

6. CONCLUSIONS

Overall, we have curated a dataset of home prices along with their street-level imagery and provided an extensive evaluation of how curb appeal affects property prices. We find that exterior appearance correlates with the price of a home and can be used to improve existing metadata-based models. Practically, these image features can be incorporated at little cost since CNN inference takes less than a second.

The results we have presented are very promising in regards to urban understanding. Applications of our work can be used for automated exterior appraisal, architectural anomaly detection, and demographic prediction. Our work opens many avenues of future research towards not only improving price estimation, but also holistically understanding urban regions.

7. REFERENCES

[1] James Q Wilson and George L Kelling, "Broken windows," *Atlantic monthly*, vol. 249, no. 3, 1982.

[2] Jelle Van Cauwenberg, Ilse De Bourdeaudhuij, Peter Clarys, Jack Nasar, Jo Salmon, Liesbet Goubert, and Benedicte Deforche, "Street characteristics preferred for transportation walking among older adults: a choice-based conjoint analysis with manipulated photographs," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 13, no. 1, 2016.

[3] Alva O. Ferdinand, Bisakha Sen, Saurabh Rahurkar, Sally Engler, and Nir Menachemi, "The relationship between built environments and physical activity: a systematic review," *American journal of public health*, vol. 102, no. 10, 2012.

[4] Stephen Malpezzi et al., "Hedonic pricing models: a selective and applied review," *Section in Housing Economics and Public Policy: Essays in Honor of Duncan Maclennan*, 2003.

[5] Gabriel Ahlfeldt, "If alonso was right: modeling accessibility and explaining the residential land gradient," *Journal of Regional Science*, vol. 51, no. 2, 2011.

[6] Genevieve Patterson and James Hays, "Sun attribute database: Discovering, annotating, and recognizing scene attributes," in *CVPR*. IEEE, 2012.

[7] Yu Su and Frédéric Jurie, "Improving image classification using semantic attributes," *IJCV*, vol. 100, no. 1, 2012.

[8] Devi Parikh and Kristen Grauman, "Relative attributes," in *ICCV*. IEEE, 2011.

[9] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva, "Learning deep features for scene recognition using places database," in *NIPS*, 2014.

[10] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Antonio Torralba, and Aude Oliva, "Places2: A large-scale database for scene understanding," *Arxiv preprint: [pending]*, 2015.

[11] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays, "Transient attributes for high-level understanding and editing of outdoor scenes," *SIGGRAPH*, vol. 33, no. 4, 2014.

[12] Daniel Glasner, Pascal Fua, Todd Zickler, and Lihi Zelnik-Manor, "Hot or not: Exploring correlations between appearance and temperature," in *ICCV*. IEEE, 2015.

[13] Sean M Arietta, Alexei Efros, Ravi Ramamoorthi, Maneesh Agrawala, et al., "City forensics: Using visual elements to predict non-visual city attributes," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 20, no. 12, 2014.

[14] Nikhil Naik, Jade Philipoom, Ramesh Raskar, and Cesar Hidalgo, "Streetscore–predicting the perceived safety of one million streetscapes," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2014.

[15] Vicente Ordonez and Tamara L Berg, "Learning high-level judgments of urban perception," in *ECCV*. Springer, 2014.

[16] Lorenzo Porzi, Samuel Rota Bulò, Bruno Lepri, and Elisa Ricci, "Predicting and understanding urban perception with convolutional neural networks," in *ACM MM*. ACM, 2015.

[17] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros, "What makes paris look like paris?," *SIGGRAPH*, vol. 31, no. 4, 2012.

[18] Stefan Lee, Nicolas Maisonneuve, David J. Crandall, Alexei A. Efros, and Josef Sivic, "Linking past to present: Discovering style in two centuries of architecture," in *IEEE International Conference on Computational Photography, ICCP*, 2015.

[19] Philip Salesses, Katja Schechtner, and César A Hidalgo, "The collaborative image of the city: mapping the inequality of urban perception," *PloS one*, vol. 8, no. 7, 2013.

[20] Daniele Quercia, Neil Keith O'Hare, and Henriette Cramer, "Aesthetic capital: what makes london look beautiful, quiet, and happy?," in *ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 2014.

[21] Fayette County Property Valuation Administrator, ," `http://www.fayette-pva.com/`.

[22] Ali S Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson, "Cnn features off-the-shelf: an astounding baseline for recognition," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2014.

[23] 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*, vol. 12, 2011.

[24] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[25] James Bergstra, Brent Komer, Chris Eliasmith, Dan Yamins, and David D Cox, "Hyperopt: a python library for model selection and hyperparameter optimization," *Computational Science and Discovery*, vol. 8, no. 1, 2015.

[26] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *IJCV*, vol. 115, no. 3, 2015.

[27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012.

[28] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *CVPR*, 2015.

[29] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.