

WEB APPENDIX TO: VISUAL PLANT IMAGERY AND PROPERTY VALUATION

WEB APPENDIX A. HEDONIC PRICING LITERATURE REVIEW

https://osf.io/sfm7k/files/4dm8p?view_only=c38f86c35600428f906cde5f7e719273

WEB APPENDIX B. REAL ESTATE AGENT SURVEY

WEB APP. B1. SURVEY QUESTIONS AND FLOW

The survey included the following questions and prompts:

Category	Details
Question - Filtering	Please tell us: In which of the following industries do you currently work? (check all that apply)" Options: real estate, property management, high technology, education, government, manufacturing, agriculture, insurance, healthcare, food service, retail or wholesale trade [Participants who did not check "Real Estate" or "Property Management" were directed to the end of the survey.]
Prompt	Introduction Thank you for participating in this study. We are interested in real estate agents' experiences and their opinions about the housing market. Please read each question carefully. There are no right or wrong answers. We are simply interested in your honest opinions. We would like to know more about the type of advice agents typically offer to individuals who wish to sell their home(s).
Question – General Advice	When a home seller is preparing to list their house on the market, what types of advice do you typically offer them? We wish to know, what are the most important things you advise home sellers preparing their homes prior to listing? Please type in the boxes below at least 3 recommendations, putting each in a different box. [10 boxes provided for open-ended responses]
Question – Presentation Advice	More specifically, to improve the presentation or appearance of a home, what are the most important recommendations for home sellers? Please type in the boxes below at least 3 recommendations, putting each in a different box. [10 boxes provided for open-ended responses]
Question – Tips Rating	How often do you recommend each of the following specific activities to home sellers who are preparing to list their house on the market? (1= Never, 2=Sometimes, 3= Always) Tips included (randomized): (1) de-clutter, (2) remove pets from the premises (for buyer visits and walk-throughs), (3) re-paint the interior in neutral tones, (4) add plants inside the home, (5) fix structural damage (e.g., roof, flooring, foundation), (6) depersonalize the space (e.g., of family photos), (7) emit a pleasant scent(s), (8) open blinds, shades, etc. to let in more light, (9) conduct a deep cleaning, (10) paint the front door, (11) add plants to the exterior, (12) touch up scuff marks, (13) fix loose handles, (14) remediate any pet odors, (15) hide valuables, (16) other (describe in box)
Question – Tips Ranking	Please rank each of the following activities a home seller should do in order to maximize the final selling price. Drag the items to indicate their importance (from 1 = most important to 15 = least important). [The list of tips was the same as in the question above.]
Questions – Agents' Experience and Demographics	Now, please tell us about yourself. Tenure: For how many years have you been working as a real estate agent? Experience: How many homes do you typically sell in one year? Price: What is the average selling price of the homes for which you act as an agent? Property type: For which type of property do you most often serve as an agent? (options included: Single family homes/Condominiums, Vacant land, Commercial, Other) Location: In which type of geographic area are most of the properties you handle located? (options included Urban, Suburban, Rural, Other) Gender: Are you: (options included: Female, Male, Other or prefer not to say) Age: How old are you?
Questions - Attention Checks	Please very briefly describe the steps involved in becoming a licensed real estate agent in your state. Please click "Disagree Completely" for this item: 1 = Disagree Completely, 5 = Agree Completely [People who didn't answer the questions properly/correctly were excluded from the analyses].

WEB APP. B2. ADDITIONAL REAL ESTATE AGENT SURVEY RESULTS

Sample: We requested 100 real estate agents from Centiment for a total price of \$995. A total of 135 participants attempted to answer the survey, among which 108 passed the filter question and completed the survey. Ten participants failed the attention check, leaving a final sample of 98 participants for analysis.

Additional results on the open-ended questions: Statements coded as referring to plants included flowers, trees, shrubs, and bushes. We observed more agents offering plant-related maintenance recommendations, such as advice on landscaping, yards, gardens, and lawns. 17.3% of agents mentioned plant-related maintenance advice when providing general guidance, and 26.5% included it when giving tips on property presentation or appearance.

TABLE B2_1. DETAILED RESULTS ON TIPS (RATING AND RANKING)

Tips	N	Rating		Ranking	
		Mean	SD	Mean	SD
De-clutter	98	2.89	.348	4.62	3.338
Conduct a deep cleaning	98	2.80	.405	5.23	3.834
Hide valuables	98	2.79	.482	8.66	4.968
Remediate any pet odors	98	2.79	.460	7.55	3.767
Remove pets from the premises (for buyer visits and walk-throughs)	98	2.74	.524	8.50	4.150
Open blinds, shades etc. to let in more light	98	2.73	.488	8.10	3.921
Fix loose handles	98	2.71	.537	8.42	3.683
Touch up scuff marks	98	2.64	.542	7.69	3.918
Emit a pleasant scent(s)	98	2.64	.561	8.73	3.713
Depersonalize the space (e.g., of family photos)	98	2.58	.641	7.39	4.148
Fix structural damage (e.g., roof, flooring, foundation)	98	2.52	.596	5.79	4.912
Re-paint the interior in neutral tones	98	2.30	.613	7.96	4.056
Paint the front door	98	2.15	.648	9.76	3.982
Add plants to the exterior	98	2.08	.620	11.00	3.650
Add plants inside the home	98	1.86	.703	11.38	3.263
Other (describe in box)	98	1.78	.780	15.21	2.703

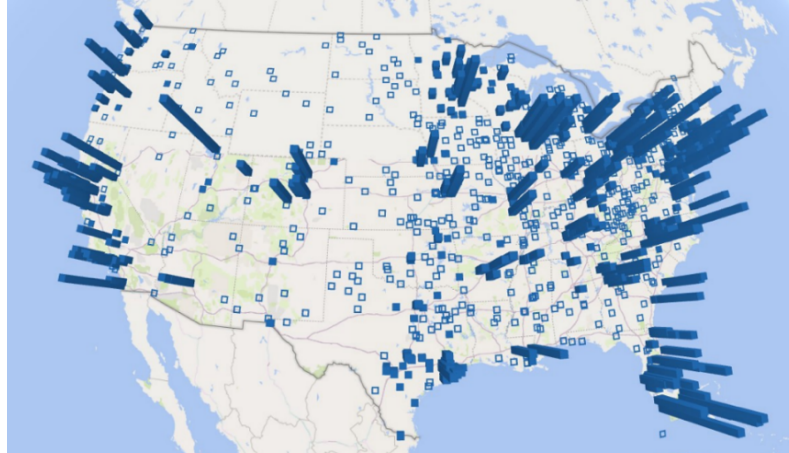
TABLE B2_2. CORRELATION BETWEEN AGENT/PROPERTY CHARACTERISTICS AND TIPS (RATING AND RANKING)

		Tips Rating (standardized) [Pearson Correlations]		Tips Ranking [Spearman's rho]	
		Interior Plants	Exterior Plants	Interior Plants	Exterior Plants
Agent's Age	correlation	.066	.155	-.120	.073
	p-value	.529	.139	.238	.472
Agent's Tenure (yrs)	correlation	.138	.082	-.082	.012
	p-value	.186	.437	.424	.904
Property Amount	correlation	.299**	.255*	.090	.001
	p-value	.004	.014	.378	.990
Property Price	correlation	.125	-.013	.023	.073
	p-value	.233	.902	.823	.475
Property Location (1 = urban, 2 = suburban, 3 = rural)	correlation	-.074	.025	-.072	.148
	p-value	.482	.812	.481	.147
N		93	93	98	98

Notes: (1) * $.01 < p < .05$, ** $.001 < p < .01$, *** $p < .001$; (2) The tip rating measure was standardized across all the 16 tips to attenuate individual rating bias. Ratings from five participants have a standard deviation of zero (i.e., they gave the same rating for all 16 tips) and were excluded from the correlation analysis.

WEB APPENDIX C. PROPERTY DISTRIBUTION AND COMPOSITION

WEB APP. C1. MAP OF 1000 U.S. ZIP CODES AND HOUSES SAMPLED



Notes: Blue squares represent the 1,000 randomly selected zip codes from all 40,933 US zip codes. Blue bars represent the number of houses sold in those zip codes during a 6-month period (6/10/19 to 12/9/19). Of the 1,000 zip codes sampled, 415 had at least one house sold; in those zip codes, the mean number of homes sold was 97.

WEB APP. C2. SAMPLED PROPERTIES VERSUS NATIONAL RECORD

	National Record	Full Sample (N = 40,294)	Analyzed Sample ¹ (N = 31,031)
Property value (\$)²			
Median property sales price	\$258,000	\$250,000	\$259,000
Average property sales price	\$384,600	\$366,232	\$367,930
Property type (house count %)³			
Single family	79.92%	76.57%	75.58%
Townhouse	9.53%	6.59%	6.65%
Condo/Co-op	10.55%	16.84%	17.76%
Region (house count %)			
Midwest	19.38%	21.42%	20.84%
Northeast	15.24%	12.30%	12.22%
South	46.49%	49.60%	48.38%
West	18.89%	16.69%	18.57%
	National Record	Our Sample (1,000 Zip Codes)	Our Sample (413 Zip Codes with Listings)
Area Type (house count %)			
Mixed/Suburban	42.73%	39.10%	30.27%
Rural	16.72%	17.60%	10.65%
Urban	40.24%	42.80%	58.60%
Unknown	0.31%	0.40%	0.48%

Notes: 1. Analyzed sample includes those with both price information and at least one image; 2. The price values (\$) of properties in our full sample are based on the 36,847 properties with available price information; 3. Only three types of properties (single-family, condominium, and townhouse) with national-level records are listed (data source: <https://www.redfin.com/us-housing-market#overview>). There is no national record of other property types available in our dataset, such as vacant land, ranch, etc.

WEB APPENDIX D. DATA SOURCES AND VARIABLE EXTRACTION PROCESSES

Data Source(s)	Procedure(s) Used to Extract Variables	Variables Extracted
Webpage Information	1. Webpage Parsing: Used "BeautifulSoup" and Regular Expression to extract information on 40,294 sold properties with 569,773 images.	<u>Outcome Variables:</u> (a) Main: Price (\$) (b) Alternatives: Days on Market, Views, Likes, Dislikes, Tours <u>Covariates:</u> (a) Property Attributes: # Bathrooms, # Bedrooms, House Size (sq ft), Lot Size (sq ft), Property Age, Property Type, Latitude, Longitude (b) Neighborhood Attributes: School Rating, Walking Score, Biking Score (c) Property Valuation Estimates: Website Estimated Price, Last Listed Price, Neighborhood Median Price <u>Fixed Effects:</u> zip codes (5-digit)
Webpage Images: plant image classifiers	1. Human Coding (5,141 randomly sampled images: 641 for model evaluation; 4,000 for supervised model training and 500 for supervised model validation): 1) Classify each image as containing plants or not; 2) Bounding boxes drawn around plant object(s) in each image 2. Algorithm Evaluation: 1) train plant image classifier for supervised models Faster-RCNN and CNNs (VGG16); 2) Evaluate the performance of different image classifiers (VOMs, VLMs, and LMMs) 3. Plant Presence/Absence Prediction: Used GPT-5-mini to predict whether each of 569,773 images contained plant life and their probabilities	<u>Independent Variables:</u> (a) Plant Image Count (b) Plant Image Percentage
Geospatial Data via Google Earth Engine ¹	1. Mask Import: Used Google Earth Engine to import mask of USGS Landsat 8 Surface Reflectance (L8SR) Tier 1 and removed noise. 2. NDVI Mapping: Generated vegetation map with bands 5 and 4 of L8SR. 3. Calculate Local Vegetation Indices: Used property coordinates to calculate NDVIs at 30, 120, and 510 m for each home.	<u>Covariates (Vegetation Indices):</u> (a) NDVI30 (b) NDVI120 (c) NDVI510
Geospatial Data via Google Map API ²	1. House Coordinate Extraction: Submitted address of each home and its neighbors to Google Map API to obtain coordinates. 2. Establish House Facing Direction: Calculated absolute direction of each house to its neighbors and bisector to establish house facing direction.	<u>Covariates:</u> (a) House Facing Direction (b) Coordinates (longitude, latitude; used for alternative geo-spatial fixed effects)
Webpage Images: CNNs ⁵	1. Human Coding: 5,121 randomly sampled images evaluated on 10 attributes: photo quality, image location, property quality, cleanliness/neatness, attractiveness, maintenance, decoration, landscaping, occupancy, and furnishing expensiveness. 2. Algorithm Training: (transfer learning and fine-tuning): 1) Architecture built for one multi-label CNN model (for photo quality, property quality, cleanliness/neatness, attractiveness, maintenance, decoration) and for four single-label CNN models (for exterior/interior, landscaping, occupancy, and furnishing expensiveness); 2) Algorithms trained with 4,000 labeled images, validated with 500 labeled images, and evaluated with 621 labeled images. 3. Image Attribute Prediction: 10 attributes of 569,773 images predicted	<u>Covariates (Image Attributes):</u> (a) Photo Quality (b) Exterior/Interior Image Count (c) Property Quality (d) Property Cleanliness/Neatness (e) Property Attractiveness (f) Property Maintenance (g) Property Decoration (h) Landscaping Quality (i) Property Occupancy (j) Furnishing Expensiveness
Webpage Images: Google Cloud Vision API + LDA	1. Label Identification: 569,773 images were submitted to Google Cloud Vision API, resulting in 3,649 textual labels for these images. 2. Label Classification: Human coders classified each of the 3,649 labels as plant-related or non-plant-related. ³ 3. Dimension Reduction: Built an LDA model to cluster the 3,138 non-plant related labels (6 clusters obtained). ⁴	<u>Covariates (Non-plant Image Content):</u> (a) Non-plant Objects in Image
Webpage Textual Descriptions: Bag-of-words, LDA	1. Word Extraction and Counts: (bag-of-words): Transformed 40,294 textual descriptions into vectors of words and calculated word counts. 2. Plant Word Calculations: (1) Human coders classified each of the 16,583 unique words (i.e., full dictionary) as plant-related or non-plant-related. ³ (2) Calculated the number and percentage of plant-related words in each textual description. 3. Topic Modeling: Used LDA to topicalize the textual descriptions (5 topics obtained) and get the topic loading scores for each home. ⁴	<u>Covariates (Textual Attributes):</u> (a) Total Word Count (b) Plant Word Count (c) Plant Word Percentage (d) Textual Topic Loadings

Notes: 1. Web App. D1. More Details on NDVI Calculation: https://osf.io/sfm7k/files/x3z6c?view_only=c38f86c35600428f906cde5f7e719273

2. Web App. D2. More Details on House Facing Direction: https://osf.io/sfm7k/files/fcu7b?view_only=c38f86c35600428f906cde5f7e719273

3. Web App. D3. Plant Word Dictionaries: https://osf.io/sfm7k/files/bvdha?view_only=c38f86c35600428f906cde5f7e719273

4. Web App. D4. More Details on LDA for Image Label Clustering and Textual Topic Modeling:

https://osf.io/sfm7k/files/zfgkd?view_only=c38f86c35600428f906cde5f7e719273

5. Web App. G. More Details on Using CNN for Image Attribute Prediction

WEB APP. E. SUMMARY OF COVARIATES AND ALTERNATIVE FACTORS

Covariates and Alternative Factors that May Influence the Results		Solutions to Addressing the Concerns
Measurement Error	Measurement Error of Plant Images Classification	i. Plant image classifier built upon GPT-5-mini, which has a classification error as low as 5% with a good balance of FPR and FNR ii. Robustness checks with alternative plant measures: plant image percentage, log-transformation of plant image count, plant image count based on other VOMs, VLMs, and LMMs Note: We also checked the results by directly quantifying the measurement error with residuals obtained from regressing plant image count on the sum of plant probabilities across all images of each property. The focal conditional association remains significant when adding the measurement error term as a covariate. There is no significant correlation between the measurement error measure and final sale price (conditional on other covariates) either.
Property and Neighborhood Attributes	Properties with some characteristics or in some neighborhoods may be more likely to have plants and sold for higher prices, e.g., larger and old properties, properties in better school districts, properties with more/less convenient transportation access	i. Conditional on property attributes: # bedrooms, # bathrooms, property size (sqft), lot size (sqft), property age, property type ii. Conditional on neighborhood attributes: school ratings, transportation scores, neighborhood median price iii. Zip code dummies or geographic group indicators as fixed effects
Property Profile Presentation	i. Sellers' behaviors and preferences may shape the presentation of home profiles, which could impact consumers' subjective perceptions ii. Properties depicting more plants in text and image may have better visual appeal on other dimensions, and signal latent property quality on dimensions that are otherwise not apparent from the property profile	i. Conditional on image attributes: non-plant image count, image quality, property quality, cleanliness/neatness, attractiveness, how well-maintained, how well-decorated, landscaping quality, property occupancy, furnishing expensiveness, non-plant image content ii. Conditional on textual attributes: word count, plant words, textual topics iii. Conditional on geospatial attributes: vegetation indices and property facing direction, which may reflect/influence the growth of plants and sellers' likelihood of using plants iv. Conditional on property valuation estimates: variables that may signal property values and sellers' expectations, including website estimated property, property listed price, neighborhood median sale price v. Robustness checks with sellers' agents fixed effects (FEs)
Sellers' Preference and Behaviors		
Market Conditions and Expectations	i. Seller expecting higher sales price may decorate their property with plants to enhance its visual appeal; ii. Owners of more expensive real estate may be more likely to incorporate plants due to lifestyles or personal preferences. iii. Real estate agents and platforms may have varying levels of private information or capacity that may influence the use of plants in property listings and final transaction price.	i. Conditional on other property valuation estimates: variables that may signal property values and sellers' expectations, including website estimated price, property listed price, neighborhood median sale price ii. Robustness checks with alternative outcome variables: price difference over the web estimated price (%) and over the listed price (%)
Others	Whether the result was driven by real-life plants or plant imagery: having more real-life plants is more likely to increase plant imagery (i.e., disentangle whether the result was driven by plant imagery or real-life plants)	i. Conditional on geospatial attributes that may reflect/influence real-life vegetation: vegetation indices, property facing direction ii. Heterogenous factors that could be influenced by plant images, but not real-life plants: interior vs. exterior images, plants shown in earlier vs. later images of a property profile iii. Robustness checks with alternative pre-sales outcome variables (likely not influenced by real-life plants): webpage views, likes, dislikes, tours
	Whether alternative conditioning factors may still exist after incorporating all covariates	i. Covariate diagnostics: specification curve, Frank's test
	Whether the OLS specification is correct	i. Robustness checks with alternative model specifications: regressions with specification curve, DML with different nuisance parameter estimation (gradient boosting, random forest, SVM, neural network), coarsened exact matching

WEB APPENDIX F. ADDITIONAL DETAILS ON PLANT IMAGE CLASSIFIERS

WEB APP. F1. MODEL SUMMARY AND COMPARISON

In order to classify each image as plant-containing or not, we compared the performance of thirteen different pre-trained or fine-tuned classification algorithms, spanning Vision-Only Models (VOMs), Vision-Language Models (VLMs), and Large Multimodal Models (LMMs). Table F1_1 summarizes the characteristics of different model categories. Table F1_2 summarizes marketing research studying unstructured image data or using LMMs.

Vision-Only Models (VOMs). Earlier research primarily relied on vision-only discriminative models, such as CNNs, for image classification. These models convert each image into matrices of RGB pixel values and map them to classification outputs through sequences of linear and nonlinear operations that extract increasingly fine-grained image features. For binary classification tasks researchers can directly use models designed for image classification to learn global image features and predict labels (e.g., VGG16; Simonyan and Zisserman 2015), or employ models designed for object detection based on regional image features (e.g., Faster R-CNN; Ren et al. 2017) and then classify each image based on whether target objects (e.g., plants) are detected. Researchers can adopt an unsupervised approach using pretrained or proprietary model architectures (e.g., Google Cloud Vision API; e.g., Li and Xie 2020), or a supervised approach customized by fine-tuning parameters on a training dataset with pre-labeled images (typically human-coded) indicating whether or not the target object is present (e.g., Liu et al. 2020). Web App. F2 provides more details on each model tested in our paper.

Vision-Language Models (VLMs). Multimodal generative machine learning models incorporate information beyond visual input to augment image analysis. VLMs, for example (e.g., CLIP and OWL-ViT) map images and text into a shared embedding space through transformers and use contrastive learning to pull matching image–text pairs closer together while pushing mismatched pairs apart (Feng et al. 2025). This enables classification through text prompts and open-vocabulary object detection. For example, CLIP jointly trains image and text encoders on large-scale image–caption pairs. Given prompts such as “an image with plants” and “an image without plants,” it compares image and text embeddings to generate probabilistic classifications (Radford et al. 2021). OWL-ViT applies the same principle to open-vocabulary object detection, locating regions that match a textual label such as “plants” and returning bounding boxes for those objects (Minderer et al. 2022, 2023). These models are typically pretrained on existing large image-language pair datasets (e.g., LAION-400M; Schuhmann et al. 2021).

Large Multimodal Models (LMMs). LMMs, such as GPT-5 (OpenAI 2025) and Gemini-2.5 (Google DeepMind 2025), incorporate an even wider range of input modalities (e.g., images, text, audio, video). By jointly reasoning across modalities, these models can follow textual instructions to extract structured information (e.g., to classify images) or generate more complex outputs (e.g., describe scenes). Most LMMs are transformer-based generative models pretrained on vast amounts of multimodal corpora from the internet and other licensed or public sources, without the need for additional training.

TABLE F1_1. IMAGE CLASSIFICATION MODELS BY CATEGORY

Model Family	Vision-Only Models (VOMs)	Vision-Language Models (VLMs)	Large Multimodal Models (LMMs)
Example Models	CNNs, Faster R-CNNs	CLIP, OWL-ViT	GPT-5, Gemini-2.5 model families
Modalities	Images only	Images + Text	Images, Text, (and sometimes audio, video)
Model Initialization			
Training Goal	Discriminative: Predict class labels or object locations in images	Generative: Learn aligned representations of images and text	Generative: General reasoning over images through textual instructions
Training Pipeline	Fully supervised	Weakly supervised (image–text alignment)	Self-supervised -> weakly supervised (multimodal alignment) -> supervised instruction tuning -> reinforcement learning via human feedback

Typical Model Size	10M – 1B parameters	50M – 5B parameters	40B – 1T+ parameters
Training Data Size	1–100M labeled images (e.g., ImageNet, COCO)	100M – 10B image–text pairs (e.g., LAION-400M)	Hundreds of billions to trillions of tokens across text, images, audios, videos
Model Application			
Zero-shot (Unsupervised)	Occasionally	Frequently	Mostly
Fine-tuning (Supervised)	Mostly	Occasionally	Rarely
Fine-tuning Dataset Requirement	~5k – 200k labeled images	~100 – 5k labeled images	~5 – 1,000 for few-shot learning ~5k – 200k+ for fine-tuning
Strengths	- Accurate on well-defined tasks, especially those with fine-grained classification/detection needs - Robust on fixed label sets with well-distributed uncertainty - Efficient inference	- Convenient zero-shot classification and open-vocabulary recognition - Flexible generalization	- Highest zero-shot performance - Strong few-shot learning capability - Highest capability in generalization (classify, describe, reason, explain)
Limitations	- Limited generalization outside trained classes - Requires labeled data most times and not zero-shot	- Less precise for well-defined tasks; not specialized for fine-grained classification or detection needs - Less robust: quality tied to textual prompt - Weaker few-shot learning capability	- Not specialized for fine-grained classification or detection tasks - Less robust: quality tied to textual prompt, may hallucinate or overly confident - Slower and more expensive

TABLE F1_2. LITERATURE REVIEW OF RELATED MARKETING RESEARCH

Topic	Literature Review
Marketing Research Using Unstructured Image Data	https://osf.io/sfm7k/files/fnpaxa?view_only=c38f86c35600428f906cde5f7e719273
Marketing Research Using Large Gen-AI Tools	https://osf.io/sfm7k/files/xv36d?view_only=c38f86c35600428f906cde5f7e719273

Performances of the image classification algorithms (Table F1_3) were assessed with a small subset of randomly sampled home images labeled as plant-containing or not by human coders. The VOMs achieved accuracies of 61.2%, 85.2%, and 90.0% (zero-shot Google Vision, and fine-tuned CNN and Faster R-CNN, respectively).¹ The VLMs achieved accuracies of 65.1% and 74.1% (CLIP, OWL-ViT2, respectively) in a zero-shot setting, indicating that incorporating textual information may improve vision task performance.² The LMMs all achieved accuracies above 91.0%. In a zero-shot setting without task-specific fine-tuning, Gemini-2.5-Flash reached accuracies of 95.0% (image-classification prompt) and 92.2% (object-detection prompt), while Gemini-2.5-Pro achieved 93.6% and 91.6%, respectively. Object detection prompting did not improve performance. The GPT-5 model family released in August 2025 delivered the strongest results: GPT-5-mini at 95.0% (image-classification prompt) and 93.3% (object-detection prompt), and GPT-5 at 95.8% and 94.7% (respectively). We adopted the GPT-5-mini algorithm to classify the images given its accuracy, efficiency, and balance between false positive and false negative rates. We used plant image count as the main independent measure, and checked robustness with alternative measures.

TABLE F1_3. PERFORMANCE OF PLANT IMAGE CLASSIFIERS

Category	Model Base	Model/Prompt Goal	Supervised/ Trained	TNs	FPS	FNS	TPs	FPR	FNR	Recall	Precision	F1 Score	Accuracy	Est. Time (days)	Est. Cost
Vision-Only Models (VOMs)	(A1) Google Vision API	Labeling	X	201	1	248	191	0.5%	56.5%	43.5%	99.5%	60.5%	61.2%	10.1	\$853
	(A2) CNN (VGG16)	Image Classification	√	170	32	63	376	15.8%	14.4%	85.6%	92.2%	88.8%	85.2%	0.2	NA

¹ The supervised approach (CNN) outperformed the unsupervised approach (Google Vision API), as it learned the distinctive visual features in our study context. The object-detection model (Faster R-CNN) further improved performance by inspecting fine-grained regional (vs. global image) features with an attentional mechanism directing the model to regions where plants were likely to appear. This reduced the number of misses (i.e., false negative rates, FNRs, of 8.4%, 14.4%, and 56.5%, respectively), particularly in interior images where plants may be small and overlooked.

² OWL-ViT2, designed for open-vocabulary object detection, outperformed CLIP, an image-classification model, particularly in reducing false positives (34.7% vs. 7.4%).

	(A3) Faster R-CNN	Object Detection	√	175	27	37	402	13.4%	8.4%	91.6%	93.7%	92.6%	90.0%	3.75	NA
Vision-Language Models (VLMs)	(B1) CLIP	Image Classification	X	132	70	154	285	34.7%	35.1%	64.9%	80.3%	71.8%	65.1%	0.2	NA
	(B2) OWL-ViT2	Object Detection	X	187	15	151	288	7.4%	34.4%	65.6%	95.0%	77.6%	74.1%	16.2	NA
Large Multimodal Models (LMMs)	(C1) Gemini-2.5-flash	Image Classification	X	185	17	15	424	8.4%	3.4%	96.6%	96.1%	96.4%	95.0%	25.0	\$275
	(C2) Gemini-2.5-flash	Object Detection	X	166	36	14	425	17.8%	3.2%	96.8%	92.2%	94.4%	92.2%	43.3	\$1,350
	(C3) Gemini-2.5-pro	Image Classification	X	163	39	2	437	19.3%	0.5%	99.5%	91.8%	95.5%	93.6%	43.3	\$2,123
	(C4) Gemini-2.5-pro	Object Detection	X	152	50	4	435	24.8%	0.9%	99.1%	89.7%	94.2%	91.6%	113.9	\$10,177
	(D1) GPT-5-mini	Image Classification	X	192	10	22	417	5.0%	5.0%	95.0%	97.7%	96.3%	95.0%	27.5	\$283
	(D2) GPT-5-mini	Object Detection	X	186	16	27	412	7.9%	6.2%	93.8%	96.3%	95.0%	93.3%	64.9	\$722
	(D3) GPT-5	Image Classification	X	195	7	20	419	3.5%	4.6%	95.4%	98.4%	96.9%	95.8%	50.1	\$1,738
	(D4) GPT-5	Object Detection	X	189	13	21	418	6.4%	4.8%	95.2%	97.0%	96.1%	94.7%	85.6	\$7,814

Notes: (1) FP = False Positive, FN = False Negative, TN = True Negative, TP = True Positive, FPR = False Positive Rate (=FPs/(FPs+TNs)), FNR = False Negative Rate (=FNs/(FNs+TPs)), Recall = TP/(TP+FNs), Precision = TP/(TP+FPs), F1 Score = 2*precision*recall/(precision + recall); (2) For CLIP, we adopted the ViT-B-32 architecture pretrained on the LAION2B-S34B-B79K dataset, using the default threshold of 50% for label detection. For OWL-ViT2, we used the checkpoint “google/owlv2-base-patch16-ensemble” with the default threshold of 10% for object detection. (3) The cost and time estimate for Google Vision API is based on information/performance as of December 2025. The time estimates for the CNN (batch size = 250) and Faster R-CNN models refer to the time required to process all 569,773 images using our fine-tuned models; model training time is not included. The time estimates for CNN, Faster R-CNN, CLIP, and OWL-ViT2 are all based on tests conducted on the researcher’s local workstation (Apple M3 Max with 96 GB of memory). The time and cost estimates for LMMs refer to the processing of all 569,773 images using real-time LMM requests as of September 2025, provided as a comparable point of reference; both time and cost can be substantially lower when using batch inference APIs.

WEB APP. F2. ADDITIONAL DETAILS ON PLANT IMAGE CLASSIFICATION WITH VOMS

F2-1. Google Vision API	https://osf.io/sfm7k/files/z726j?view_only=c38f86c35600428f906cde5f7e719273
F2-2. CNN (VGG16)	https://osf.io/sfm7k/files/48mzt?view_only=c38f86c35600428f906cde5f7e719273
F2-3. Faster R-CNN	https://osf.io/sfm7k/files/ry6fg?view_only=c38f86c35600428f906cde5f7e719273

WEB APP. F3. LMM PROMPTS FOR PLANT IMAGE CLASSIFICATION

Prompt Goal	System Prompt	User Prompt	Output Structure
Image Classification	"You are an assistant for image classification."	"Please classify the image based on whether it contains plants or not. Respond in structured JSON that matches the schema fields: plant_label, plant_probability."	Field 1 - "plant_label" (integer 0 or 1): "1 if the image contains plants, 0 if not" Field 2 - "plant_probability" (number 0-100): "Probability (0–100%) that the image contains plants"
Object Detection	"You are an assistant for object detection."	"Please detect all plant objects in the image and classify the image based on whether it contains plants or not. Respond in structured JSON that matches the schema fields: plant_label, plant_locations. For plant_locations, it should be a list of bounding boxes, where each box (2D) is in the format of [xmin,ymin, xmax, ymax] and each coordinate value is scaled to 0-100 as a percentage of the image size. Never return masks or code fencing. Limit to a maximum of 30 plant objects."	Field 1 - "plant_label" (integer 0 or 1): "1 if the image contains plants, 0 if not" Field 2 - "plant_locations" (list, length 0-30): "locations of plant objects detected as a list of bounding boxes, where each box (2D) is in the format of [xmin,ymin, xmax, ymax] and each coordinate value is scaled to 0-100 as a percentage of the image dimensions" Sub-attribute of Field 2 (array, length=4): "bounding box coordinates in the format of [xmin,ymin, xmax, ymax]" Sub-attribute of sub-attribute (number 0-100): "coordinate value scaled to 0-100 as a percentage of the image size"

WEB APPENDIX G. CNN FOR OTHER IMAGE ATTRIBUTES

To evaluate the image attributes (see Table G_1 for a summary of all the image attribute variables) for all property images, we trained five separate CNNs built upon VGG16 (Simonyan and Zisserman 2015). The CNN architectures are similar to those used for plant image classification (see Web App. F2-2), except that the last dense layers are configured with different activation functions and node amounts, so that the output layer can conform to single or multiple continuous or binary classification outcomes accordingly. Table G_2 summarizes the model architecture and the learning process.

TABLE G_1. MEASURES OF IMAGE ATTRIBUTES AND CNN MODELS

	Measure	Coding Questions for the Training Dataset	Prediction Outcome	CNN Model	Mean Squared Error	Mean Absolute Error
1	Photo Quality	How would you rate the overall quality of the photo? (0 = very low, 4 = very high)	0-1 Continuous Value	MCNN1	0.043	0.157
2	Property Quality	How would you rate the overall quality of the property depicted in the image? (0 = very low, 4 = very high)	0-1 Continuous Value	MCNN1	0.039	0.154
3	Cleaness/Neatness	How clean/neat do you think the property depicted is? (0 = dirty/messy, 4 = clean/neat)	0-1 Continuous Value	MCNN1	0.040	0.152
4	Property Attractiveness	How attractive do you think the property depicted is? (0 = unattractive, 4 = attractive)	0-1 Continuous Value	MCNN1	0.044	0.164
5	Property Maintenance	How well-maintained do you think the property depicted is? (0 = not well maintained, 4 = well-maintained)	0-1 Continuous Value	MCNN1	0.039	0.155
6	Property Decoration	How well-decorated do you think the property depicted is? (0 = not well decorated, 4 = well-decorated)	0-1 Continuous Value	MCNN1	0.046	0.168
7	Exterior/Interior	Is this image taken interior or exterior of the property? (0 = interior, 1 = exterior)	0-1 Categorical Value	MCNN2	0.019	0.019
8	Landscaping Quality	If the image is taken exterior of the property, what do you think the quality of the landscaping is? (0 = low quality landscaping, 4 = high quality landscaping)	0-1 Continuous Value	MCNN3	0.065	0.207
9	Property Occupancy	If the image is taken interior of the property, how occupied you think the property is? (0 = vacant, 4 = occupied)	0-1 Continuous Value	MCNN 4	0.104	0.265
10	Furnishing Expensiveness	If the property is not vacant, how expensively furnished you think the property is? (0 = inexpensively furnished, 4 = expensively furnished)	0-1 Continuous Value	MCNN 5	0.055	0.200

TABLE G_2. VGG16 MODEL ARCHITECTURE FOR CLASSIFICATION AND EVALUATION

Layer		Layer Type	Output Dimensions	Weight Layers	Trainable	Learning Method (Learning Rate)
Bottom Layers	1	Input Layer	$224 \times 224 \times 3$	No	No	
	2	Convolutional Layer	$224 \times 224 \times 64$	Yes	No	Transfer Learning
	3	Convolutional Layer	$224 \times 224 \times 64$	Yes	No	Transfer Learning
	4	Max Pooling Layer	$112 \times 112 \times 128$	No	No	Max Pooling
	5	Convolutional Layer	$112 \times 112 \times 128$	Yes	No	Transfer Learning
	6	Convolutional Layer	$112 \times 112 \times 128$	Yes	No	Transfer Learning
	7	Max Pooling Layer	$56 \times 56 \times 256$	No	No	Max Pooling
	8	Convolutional Layer	$56 \times 56 \times 256$	Yes	No	Transfer Learning
	9	Convolutional Layer	$56 \times 56 \times 256$	Yes	No	Transfer Learning
	10	Convolutional Layer	$56 \times 56 \times 256$	Yes	No	Transfer Learning
	11	Max Pooling Layer	$28 \times 28 \times 512$	No	No	
	12	Convolutional Layer	$28 \times 28 \times 512$	Yes	No	Transfer Learning
	13	Convolutional Layer	$28 \times 28 \times 512$	Yes	No	Transfer Learning
	14	Convolutional Layer	$28 \times 28 \times 512$	Yes	No	Transfer Learning
	15	Max Pooling Layer	$14 \times 14 \times 512$	No	No	Max Pooling
	16	Convolutional Layer	$14 \times 14 \times 512$	Yes	Yes	Fine-tuning (0.005)
	17	Convolutional Layer	$14 \times 14 \times 512$	Yes	Yes	Fine-tuning (0.005)
	18	Convolutional Layer	$14 \times 14 \times 512$	Yes	Yes	Fine-tuning (0.005)
	19	Max Pooling Layer	$7 \times 7 \times 512$	No	Yes	
Top Layers (Fully Connected Layers)	20	Flatten Layer	25088	No	Yes	Max Pooling
	21	Dense Layer (ReLU)	4096	Yes	Yes	(0.005)
	22	Dense Layer (ReLU)	1024	Yes	Yes	(0.005)
	23	Dropout Layer	512	No	Yes	Dropout = 50%
	24	Dense Layer (Sigmoid)	Model-dependent	Yes	Yes	(0.005)
Output Layer	25	Output layer	Model-dependent	No	Yes	0-1 continuous measures

Notes: The computer first reshapes each property image, transforms it into a collection of pixels with RGB values in a matrix form, and uses this information as the input layer. Then 18 bottom layers (including 13 convoluted layers alternated by five max pooling layers) extract the most basic image features

(accounted by the 13 convolutional layers), such as brightness, color, contrast, curves, edges, circle, and strips (Zeiler and Fergus 2014, Zhang and Luo 2023), and deal with dimension reduction and ensure a certain degree of shift invariance (accounted by the four max pooling layers alternated by convolutional layers; (Scherer et al. 2010)). The subsequent five fully connected top layers flatten the multi-dimensional output from the bottom layers and extract features more critical to plant identification. Specifically, the last dense layer maps the remaining 512 feature nodes into a number with an activation function, "sigmoid" (mathematical representation is: $S(x) = \frac{1}{1+e^{-x}}$), which is similar to the logistic function that restrict the output to a range of 0 to 1.

To build the training dataset, four human coders evaluated the 5,121 randomly sampled house images on the ten target features, with 4,000 images used for model training, 500 for model validation, and 621 for model prediction or testing. Each coder coded approximately 1,280 images. To de-bias individual differences in image ratings (e.g., some raters may give higher average ratings than others), we first did mean adjustment. For the nine 5-point scale items (i.e., photo quality, property quality, cleanliness/neatness, attractiveness, maintenance, decoration, landscaping quality, occupancy, furnishing expensiveness), the rating scores from each coder were scaled by multiplying the overall mean value for that item across all 5,121 images, and dividing by the mean value for that item among the images rated by that coder. After the mean adjustment, each measure was scaled to 0-1 by dividing the highest rating value on that measure.

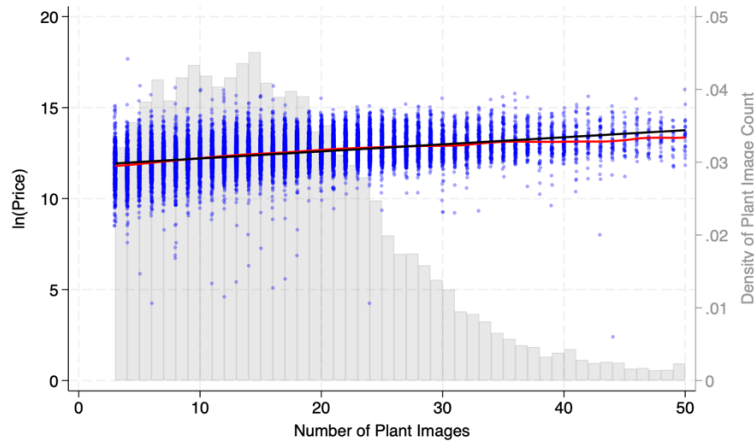
The first CNN model (MCNN1) was constructed with an output space of six dimensions to predict the six continuous variables that applied to all images: photo quality, property quality, cleanliness/neatness, attractiveness, maintenance, and decoration (multi-label CNN, Liu et al. 2020). We used "sigmoid" as the activation function for the last layer (layer 24) to restrict the output to a range of 0 to 1. The model was again trained with transfer learning (layers 2 – 15 as in Table G_2) and fine-tuning (layers 16 - 19 as in Table G_2, the learning rate was 0.005). The training process aims to minimize the mean squared error between the predicted and human-coded outcomes.

The second model (MCNN2) we constructed was to predict whether each image is an interior or exterior image. The model architecture was the same as in Table G_2. Given that parameters in Model 1 had already captured some fundamental visual information of our image dataset, we used the same parameters of all the bottom layers (i.e., layers 2-19 as in Table G_2) as that of the tuned Model 1. The training process only tuned parameters at the top layers (i.e., layers 20-24 as in Table G_2) to minimize the misclassification error. The model prediction accuracy reached 98.1%.

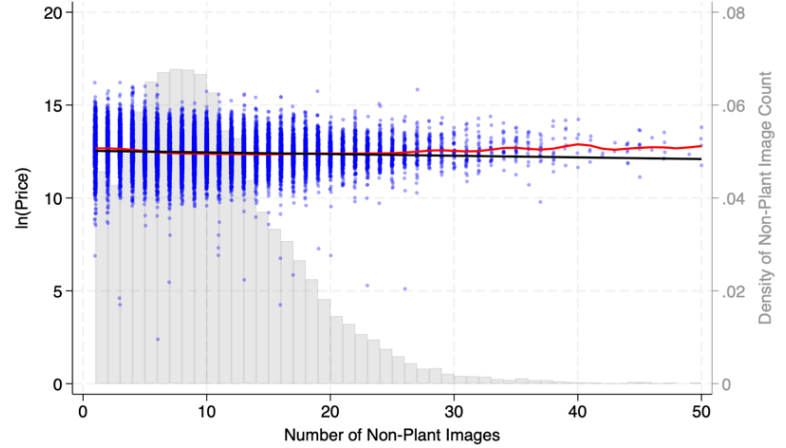
The remaining three variables (i.e., landscape quality, occupancy, and furnishing expensiveness) conditioning on exterior/interior images were evaluated by three separate CNN models (MCNN 3-5). The model architecture and activation functions for these three models were similar to that of MCNN1, except that the output dimension was set as one (i.e., such that one model predicts one variable). Similar to MCNN2, only the top layers were trainable. Similarly, the training goal is to minimize the mean squared error between the predicted and human-rated measures. Compared to MCNN1, these three models have relatively low performance, partly due to the fact that the three variables are conditioned on whether an image is interior or exterior. With much smaller training samples, it is reasonable to expect lower accuracy.

WEB APPENDIX H. RELATIONSHIP BETWEEN PRICE AND IMAGE COUNT

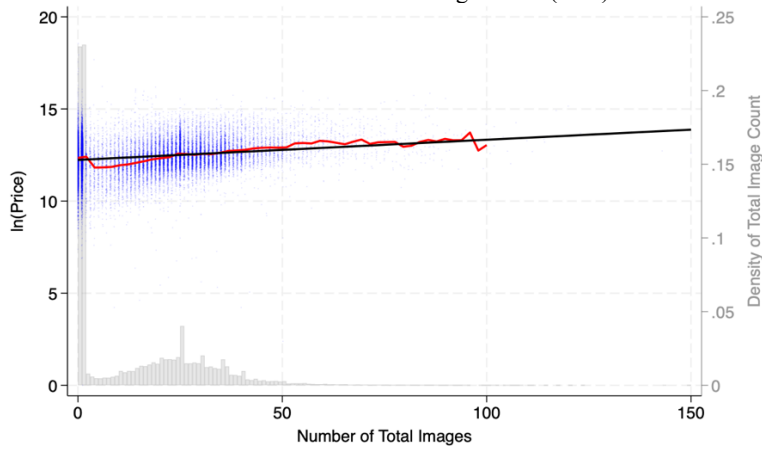
H1. Price and Plant Image Count (Zoomed-in)



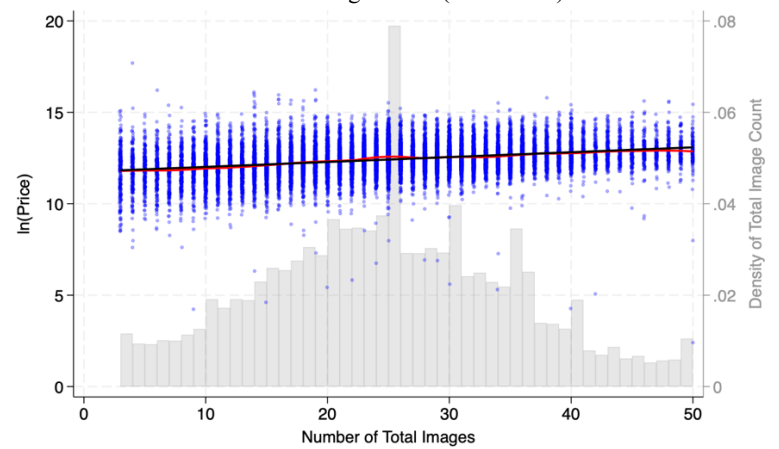
H2. Price and Non-Plant Image Count (Zoomed-in)



H3. Price and Total Image Count (Raw)



H4. Price and Total Image Count (Zoomed-in)



Notes: The scatterplot has jitter added to the x-axis, as the numbers of images are integers. The nonlinear fitted lines used kernel-weighted local polynomial smoothing with the default Epanechnikov kernel and rule-of-thumb optimal bandwidth settings in Stata.

WEB APP. I. MAIN REGRESSION RESULTS
WEB APP. I1. MAIN REGRESSION TABLE 2 CONTINUED

	(1) Price(ln)	(2) Price(ln)	(3) Price(ln)	(4) Price(ln)	(5) Price(ln)
Additional Property Attributes:					
Lot Size (missing)		0.2459 ⁺ (0.1265)	0.2851*** (0.1059)	0.3303*** (0.0604)	0.0003 (0.0139)
Property Style (0 = Condo)					
1 = Mobile		-0.3821*** (0.1137)	-0.3129*** (0.1025)	-0.2900*** (0.0841)	-0.0472* (0.0210)
2 = Multi-family		-0.0671 (0.0962)	0.0657 (0.0911)	0.2878*** (0.0356)	-0.0063 (0.0092)
3 = Other		-0.0283 (0.0660)	0.0349 (0.0580)	0.2036*** (0.0407)	-0.0418*** (0.0079)
4 = Parking		-0.0532 (0.6582)	0.1037 (0.6464)	0.3519*** (0.0877)	0.0078 (0.0113)
5 = Ranch		0.3747*** (0.1005)	0.3938*** (0.0903)	0.5224*** (0.0785)	0.0441*** (0.0165)
6 = Single Family		0.2348*** (0.0622)	0.2563*** (0.0519)	0.3760*** (0.0312)	-0.0324*** (0.0059)
7 = Timeshare		-0.9479*** (0.1067)	-1.1462*** (0.1101)	-2.0903*** (0.0486)	
8 = Townhouse		0.0313 (0.0936)	0.0422 (0.0897)	0.2337*** (0.0303)	-0.0117* (0.0058)
9 = Land		0.0266 (0.0770)	0.1322 ⁺ (0.0690)	0.2859*** (0.0466)	-0.0279*** (0.0081)
House Facing Direction (baseline = -180° to -135°, south to southeast)					
-135° to -90° (southeast to east)		0.0432 ⁺ (0.0239)	0.0356 ⁺ (0.0214)	0.0017 (0.0136)	0.0005 (0.0095)
-90° to -45° (east to northeast)		0.0135 (0.0223)	0.0111 (0.0198)	0.0097 (0.0136)	0.0077 (0.0094)
-45° to 0° (northeast to north)		0.0241 (0.0201)	0.0268 (0.0174)	0.0117 (0.0123)	0.0039 (0.0091)
0° to 45° (north to northwest)		0.0028 (0.0200)	0.0032 (0.0181)	0.0005 (0.0121)	0.0058 (0.0090)
45° to 90° (northwest to west)		0.0105 (0.0208)	0.0105 (0.0188)	0.0069 (0.0142)	0.0054 (0.0094)
90° to 135° (west to southwest)		0.0193 (0.0220)	0.0219 (0.0200)	0.0055 (0.0143)	-0.0006 (0.0106)
135° to 180° (southwest to south)		0.0162 (0.0242)	0.0127 (0.0225)	0.0122 (0.0160)	0.0014 (0.0110)
Additional Image Attributes					
Non-plant Objects					
Cluster1			0.8012*** (0.1482)	0.2298*** (0.0864)	0.0603 ⁺ (0.0353)
Cluster2			0.7639*** (0.1488)	0.3016*** (0.0767)	-0.0216 (0.0268)
Cluster3			0.5991*** (0.1285)	0.0607 (0.0611)	-0.0174 (0.0264)
Cluster4			0.7252*** (0.1682)	0.2057* (0.0871)	-0.0188 (0.0265)
Cluster5			0.5885*** (0.1184)	0.1412* (0.0615)	-0.0382* (0.0216)
Additional Textual Attributes					
Textual Topics					
Topic1			0.2339 (0.2664)	0.076 (0.1200)	0.1756*** (0.0465)
Topic2			1.8040*** (0.3157)	0.9733*** (0.1523)	0.0687 (0.0519)
Topic3			0.2532 (0.2566)	0.3059* (0.1419)	0.3068*** (0.0554)
Topic4			-1.0448*** (0.2858)	-0.2132* (0.1221)	0.1103* (0.0462)
Fixed Effects	No	No	No	zip code	zip code
Property/Neighborhood Attributes	No	Yes	Yes	Yes	Yes
Image/Text Attributes	No	Yes	Yes	Yes	Yes
Number of FE Groups	NA	NA	NA	319	309
Observations	31031	25607	25607	25590	22409
Adjusted R-sqr	0.1002	0.4831	0.5514		
Within-cluster R-sqr				0.5378	0.8675

Notes: (1) ⁺ 0.05 < *p* < 0.10, * 0.01 < *p* < 0.05, ** 0.001 < *p* < 0.01, *** *p* < 0.001; (2) values without parentheses are coefficient estimates, and values with parentheses are standard errors of coefficient estimates; (3) The model 2 results show that newer houses and larger houses, with more rooms and more square footage, tended to sell for more (as might be expected), helping to control for the opportunity to include more plant images. The results also suggest that houses with better transportation and education features tend to sell at higher prices. The vegetation indexes had negative coefficients, possibly reflecting less vegetation in more expensive, densely populated neighborhoods. (4) In model 3, exterior image counts had a negative coefficient, possibly because interior images tended to provide more information than exterior images. Most of the other image attributes had positive coefficients (except for photo quality, likely because this information was captured by other image variables). Including these variables individually had a similar impact on the coefficient for plant images as including them as a group, suggesting that the image variables are adequately accounting for visible qualities that correlate with plant images. Since the image attributes were highly correlated with each other, we conducted a series of regressions that included these variables in Model 5 one at a time to assess their impact on the main model. This resulted in significantly positive estimates for each of the image attributes, with point estimates in the range of .017 to .019 and smaller effects for landscaping (.004) and occupied (.003). Although the six non-plant image content clusters added explanatory power to the model, testing them incrementally had little impact on the coefficient for plant images. Model 3 also added textual attributes. Longer text descriptions and more plant-related words, which may imply more marketing effort and better living environments, were linked to higher prices.

WEB APP. I2. SPECIFICATION CURVE

We used specification curve (Simonsohn et al. 2020) to display the estimation for the association between property sales price (log-transformed) and plant or non-plant image count across 64 non-redundant linear models in one graph, with the conditioning factors operationalized as all possible combinations of the seven variable sets (in the order as they appear in Figures I2_1, I2_2: property attributes, neighborhood attributes, image variables, textual attributes, website estimated price, last listed price, and zip code fixed effects; all specifications included zip code fixed effects). This allows us to examine whether the estimation for the focal coefficient of interest is robust and not susceptible to covariate selection bias. Theoretically, once enough covariates are included, the estimates would vary within a narrow range and not statistically differ from each other.³ Figure I2_1 displays the results using common samples across all the 64 modes, and Figure I2_2 displays the results using model-specific samples.

FIGURE I2_1. SPECIFICATION CURVE USING COMMON SAMPLES

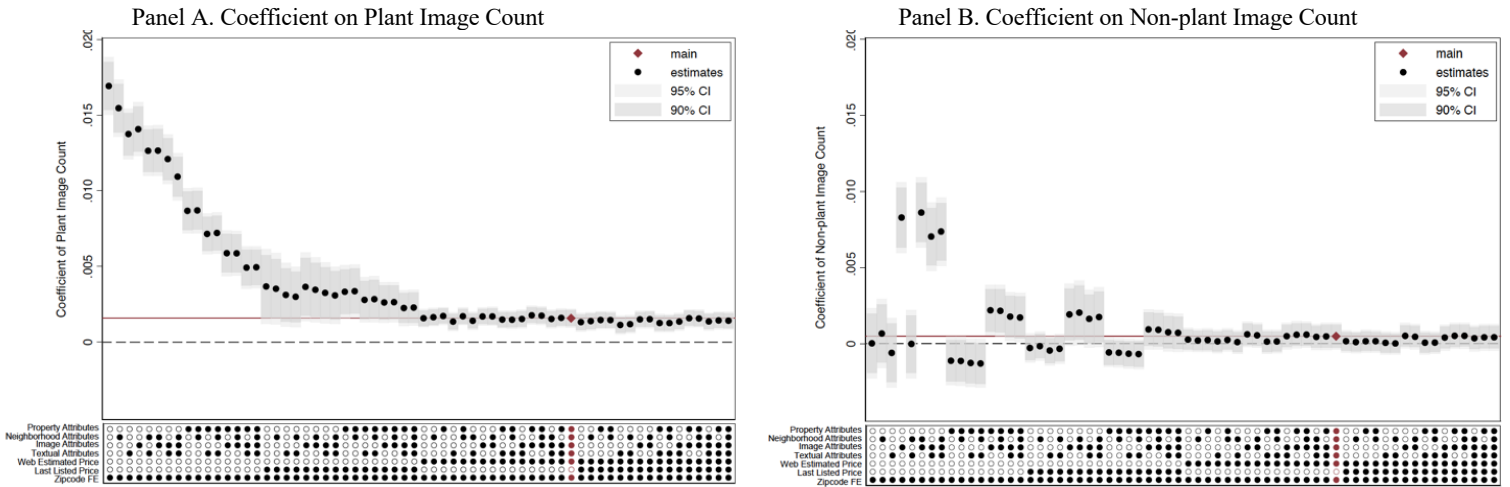
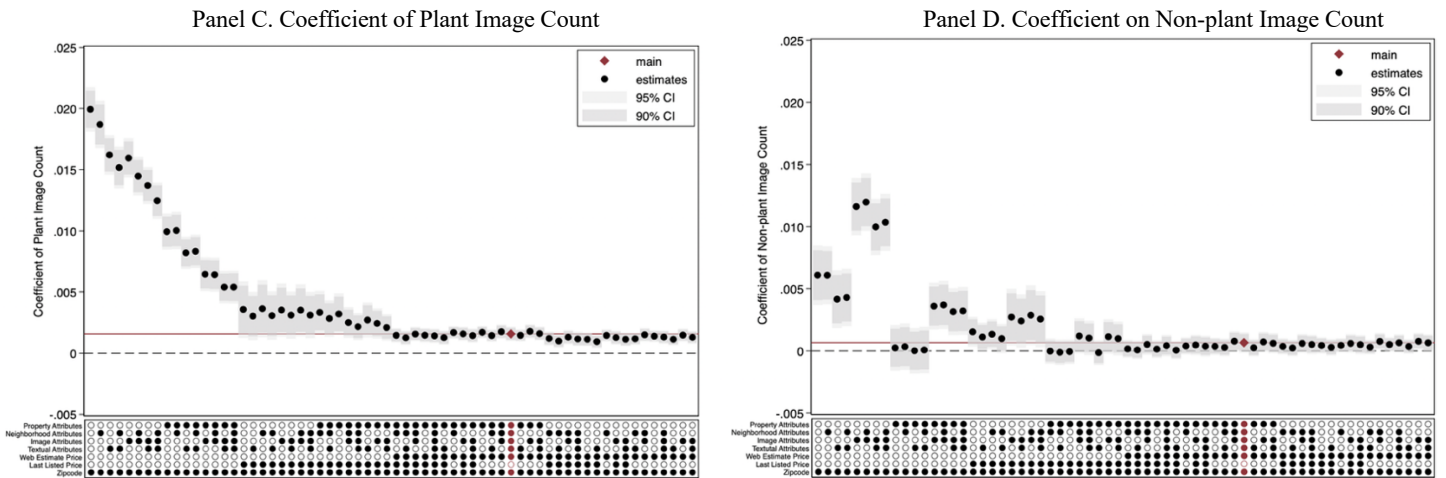


FIGURE I2_2. SPECIFICATION CURVE USING MODEL-SPECIFIC SAMPLES



³ The results have low sensitivity to additional covariates once enough covariates are added to the model, suggesting that the group of covariates we obtained are likely collectively capturing the highest-impact covariates.

WEB APPENDIX J. ADDITIONAL RESULTS

WEB APP. J1. RESULTS WITH ALTERNATIVE PLANT IMAGERY MEASURES

Image Classifier	GPT-5-mini		Google	CNN	Faster R-CNN	CLIP	OWL-ViT2
Plant Imagery Measure	(6) Plant Image %	(7) ln (Plant Image Count + 1)	(8) Plant Image Count	(9) Plant Image Count	(10) Plant Image Count	(11) Plant Image Count	(12) Plant Image Count
Coefficient of Plant Imagery	0.0407*** (0.0096)	0.0222*** (0.0028)	0.0010 (0.0006)	0.0015*** (0.0003)	0.0016*** (0.0003)	0.0066*** (0.0013)	0.0013*** (0.0004)
Coefficient of Non-plant Imagery		0.0030 (0.0028)	0.0013*** (0.0003)	0.0010** (0.0003)	0.0008+ (0.0004)	0.0011*** (0.0002)	0.0012*** (0.0003)
Coefficient Comparison	-	$p = .000$	$p = .645$	$p = .253$	$p = .086$	$p = .000$	$p = .829$

Note: (1) For all the models, we used the log-transformed price as the dependent variable, and used the same set of covariates as our Model (5); (2) Values without parentheses are coefficient estimates, and values with parentheses are standard errors of coefficient estimates; (3) + $0.05 < p < 0.10$, * $0.01 < p < 0.05$, ** $0.001 < p < 0.01$, *** $p < 0.001$. (4) Model 7 used the log-transformation of the Non-plant Image Count as a covariate.

WEB APP. J2. RESULTS ON ALTERNATIVE MARKET PERFORMANCE MEASURES

Market Performance Measure	(18) Price (\$)	(19) Price Difference Over Web Estimated Price (%)	(20) Price Difference Over Listed Price (%)	(21) Price Difference Over Neighborhood Median Price (%)
Coefficient of Plant Image Count	934.7*** (300.7)	0.0010*** (0.0001)	0.0003+ (0.0001)	0.0016*** (0.0004)
Coefficient of Non-plant Image Count	-992.2*** (220.3)	0.0002 (0.0001)	0.0000 (0.0002)	0.0000 (0.0004)
Coefficient Comparison	$p = .000$	$p = .000$	$p = .082$	$p = .004$

Note: (1) For all the models, we used the same set of covariates as our preferred Model 5; (2) Values without parentheses are coefficient estimates, and values with parentheses are standard errors of coefficient estimates; (3) + $0.05 < p < 0.10$, * $0.01 < p < 0.05$, ** $0.001 < p < 0.01$, *** $p < 0.001$; (4) Model 18 excluded houses with extreme prices (top and bottom 1%), and Model 19-21 excluded houses with extreme price difference over the benchmark price (beyond a 1:2 ratio, i.e., -50% or +100% difference).

WEB APP. J3. RESULTS WITH ALTERNATIVE PROPERTY VALUE ESTIMATES

Price Control	(22) Last Listed Price (ln)	(23) Neighborhood Median Price (ln)	(24) Last Listed Price (ln) + Neighborhood Median Price (ln)
Coefficient of Plant Image Count	0.0022*** (0.0006)	0.0053*** (0.0007)	0.0022*** (0.0006)
Coefficient of Non-plant Image Count	0.0010 (0.0007)	0.0030** (0.0010)	0.0013* (0.0007)
Coefficient Comparison	$p = .066$	$p = .011$	$p = .106$

Note: (1) For all the models, we used the log-transformed price as the dependent variable, and used the same set of covariates as our Model (5); (2) Values without parentheses are coefficient estimates, and values with parentheses are standard errors of coefficient estimates; (3) + $0.05 < p < 0.10$, * $0.01 < p < 0.05$, ** $0.001 < p < 0.01$, *** $p < 0.001$.

WEB APP. J4. RESULTS WITH ALTERNATIVE FIXED EFFECTS

Our main model included five-digit zip codes as dummies, which allows us to estimate the conditional association between plant image and property transaction price by comparing outcomes within geographic clusters, in the same spirit of fixed effects (FE) models. To the extent that treating the 5-digit zip code as a dummy variable may obscure certain geographic patterns embedded in the numeric structure of the zip code, we also examined the results using alternative geographic clusters.

Model 25 (Table J4_1) used more coarse-grained three-digit zip code as dummies. For Models 26 through 29 (Table J4_1), we constructed geographic groupings grids based on coordinates. We varied the size of the grid cells by using 1, 2, 3, or 4 decimal places of the latitude/longitude coordinates, which roughly corresponded to 10,000, 1000, 100, 10 meters, respectively. The results held across all alternative geographic cluster measures (note that the sample size became progressively smaller due to lack of matches in the finer grids). We also checked the results by adding sellers' agent as dummies to our main model (5) and obtained consistent result (Model 30, Table J4_1). Because this variable is noisy with

many missing values⁴, we did not include it in our main analysis.

TABLE J4_1. ALTERNATIVE FIXED EFFECTS

Alternative FE	(25) Zip-3	Latitude/longitude-based Geographic Grids				(30) Zip5 + Seller's Realtor
		(26) LL-1d	(27) LL-2d	(28) LL-3d	(29) LL-4d	
Coefficient of Plant Image Count	0.0015*** (0.0003)	0.0016*** (0.0003)	0.0016*** (0.0003)	0.0022*** (0.0004)	0.0031*** (0.0006)	.0017*** (0.0003)
Coefficient of Non-plant Image Count	0.0007 (0.0004)	0.0006 (0.0004)	0.0007+ (0.0004)	0.0016* (0.0006)	0.0017* (0.0007)	0.0011** (0.0004)
Coefficient Comparison	$p = .085$	$p = .036$	$p = .087$	$p = .330$	$p = .015$	$p = .313$

Note: (1) For all the models, we used the same set of covariates as our preferred Model 5; (2) Values without parentheses are coefficient estimates, and values with parentheses are standard errors of coefficient estimates; (3) + $0.05 < p < 0.10$, * $0.01 < p < 0.05$, ** $0.001 < p < 0.01$, *** $p < 0.001$; (4) zip5 represents 5-digit zip code, zip3 or zip-3 represents the first three digits of zip code; LL-1d, LL-2d, LL-3d, LL-4d represent fixed effects of grid cells by using 1, 2, 3, or 4 decimal places of the coordinates.

WEB APP. J5. RESULTS BY IMAGE CHARACTERISTICS

(31) Interior vs. Exterior Plant Images		(32) Plants in First 10 vs. Later Images		(33) Plants in First 20 vs. Later Images	
Coefficient of Interior Plant Percentage (%)	0.0448*** (0.0117)	Coefficient of Plant Image Count (first 10)	0.0034*** (0.0010)	Coefficient of Plant Image Count (first 20)	0.0027*** (0.0007)
Coefficient of Exterior Plant Percentage (%)	0.0090 (0.0113)	Coefficient of Plant Image Count (excluding first 10)	0.0013** (0.0004)	Coefficient of Plant Image Count (excluding first 20)	0.0008+ (0.0005)
Coefficient Comparison	$p = .028$	Coefficient Comparison	$p = .102$	Coefficient Comparison	$p = .026$

Note: (1) For all the models, we used the log-transformed price as the dependent variable with the same set of covariates as our Model (5); (2) Values in parentheses are standard errors; (3) + $0.05 < p < 0.10$, * $0.01 < p < 0.05$, ** $0.001 < p < 0.01$, *** $p < 0.001$.

⁴ In our dataset, only 26,072 out of 31,031 properties have realtor information. In total, we identified 6871 unique realtors; each realtor was associated with 34 properties on average, and 3790 realtors handled only one property in our dataset.

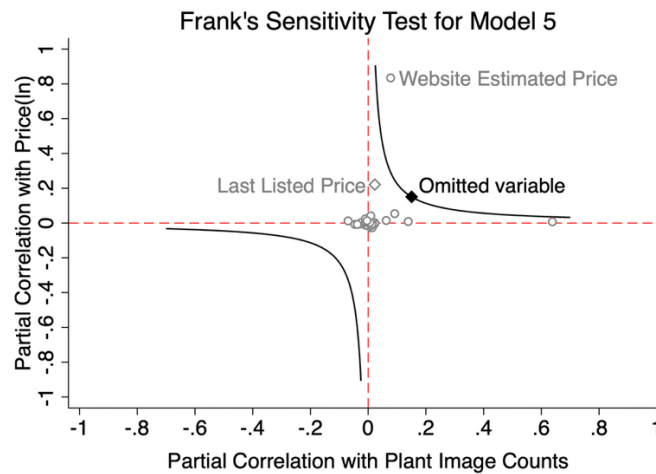
WEB APP. K. ADDITIONAL ANALYSES IN GENERAL DISCUSSION

WEB APP. K1. FRANK'S TEST

This paper aims to investigate an important factor in house valuation, i.e., plant imagery, beyond alternatives that may correlate with plant imagery or property sales price. Thus, we carefully curated an extensive set of covariates to account for conceivable alternatives as comprehensively as possible. To test the sufficiency of the covariate set, we used Frank's Test to assess the degree to which a hypothetical variable would need to be correlated with the key variable of interest (property price and plant image count) to invalidate our inference of the conditional association.

Frank's test computes the minimum product of the partial correlations between the hypothetical conditioning variable and both the focal independent and dependent variables that could render our estimate no longer statistically significant when adding the hypothetical variable to a model, i.e., ITCV (impact threshold of a confounding variable).⁵ For the model with the full set of covariates included (Model 5), $\sqrt{ITCV} = .151$, meaning an alternative variable that would render the targeted conditional association no longer significant at the 5% level should have a partial correlation of over .151 with both price and plant image counts. The only candidate variable in our data with partial correlations of that magnitude is the real estate website's estimated valuation (see Figure K1_1 for visualization). In other words, even after accounting for our strongest control, the web estimated value, there would need to be another variable with as strong partial correlations with plant image count and property price as the web estimated value to invalidate our inference on the conditional association, which is unlikely given that the website claims the estimated price has an error rate under 2% of the final sale price. Thus, the group of covariates curated, including the property, neighborhood, image, and text attributes, as well as the zip code fixed effects and property estimated valuation, are likely effective in collectively accounting for almost all conceivable alternatives.

FIGURE K1_1. FRANK'S TEST FOR MODEL 5



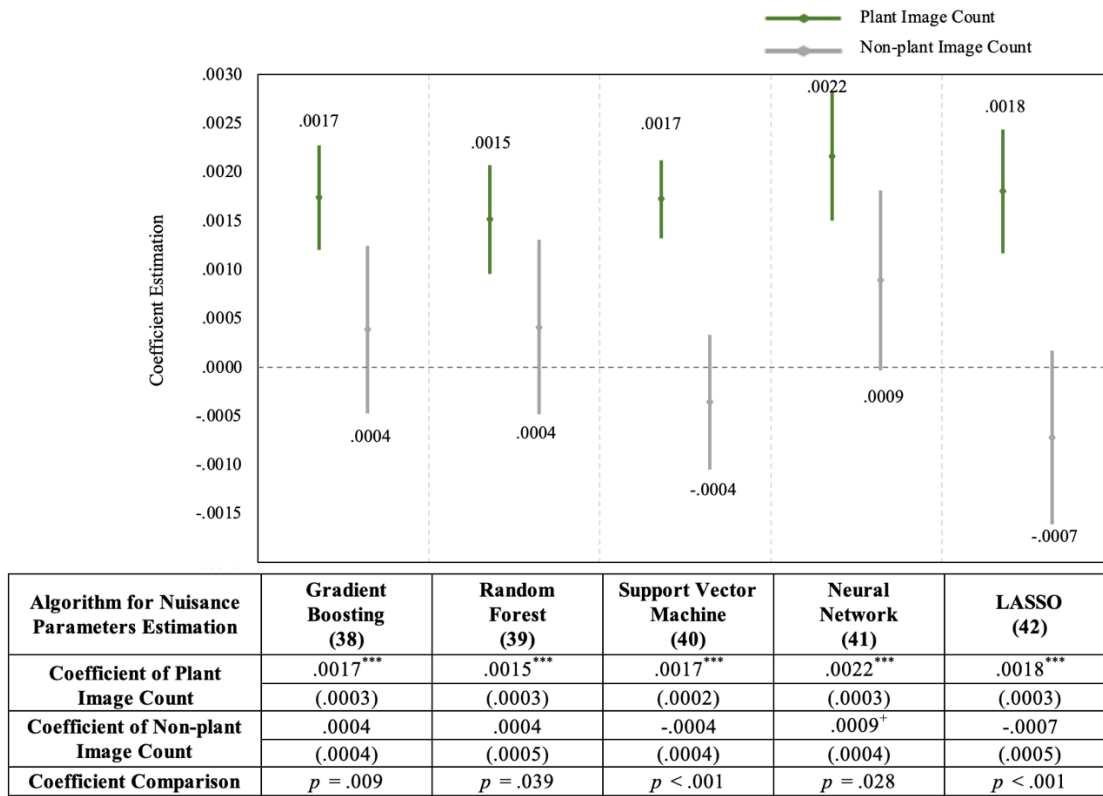
Note: (1) Curves depict partial correlations between a hypothetical covariate and price and plant image counts that meet ITCV and may potentially cause the 95% confidence interval of the conditional association estimate in Model 5 to include 0. (2) Points marked with a circle represent variables included in model 5, and points marked with diamonds represent variables not yet included in the model 5. (3) The point labeled with 'hypothetical covariate' represents partial correlation pairs equal to the square root of the ITCV, which is the lowest-impact hypothetical covariate that could invalidate the conditional association estimate.

⁵ See Frank (2000) for the derivations, and see Busenbark et al. (2022) for a discussion of its application. We calculated ITCV with the 'konfound' command in Stata (Xu et al. 2019). The plot was generated by computing the partial correlation between each covariate and price or plant image count one by one, following Torrats-Espinosà (2021).

WEB APP. K2. DOUBLE MACHINE LEARNING

We also ran double machine learning (DML; Chernozhukov et al. 2017, see Knaus 2021 for an application) to check the relationship between plant image count and property transaction price, after partialling out potential influences from the covariates (as used in the main model 5) with non-parametric machine-learning-based algorithms. Using Gradient Boosting for the first-stage nuisance parameter estimation (Model 38), the DML estimation produced a positive and significant coefficient for plant image count ($\beta = .0017$, $SE = .0003$, $p < .001$), but not for non-plant image count ($\beta = .0003$, $SE = .0004$, $p = .481$). The coefficient for plant images is significantly larger than that for non-plant images ($p = .008$), consistent with our main findings. Similar results were observed using Random Forest (Model 39), Support Vector Machine (Model 40), Neural Network (Model 41), and LASSO (Model 42) as the first-stage estimation approach (Web App. Figure K2_1).

FIGURE K2_1. DML ESTIMATES



WEB APP. K3. COARSENEDED EXACT MATCHING

We also checked the relationship between plant image count and property transaction price using an alternative approach: *coarsened exact matching*, wherein we re-estimated our models by comparing within groups based on both geography and other covariates (Iacus et al. 2012). This method allows researchers to ensure covariate balance, reducing the reliance on parametric assumptions for control variables while maintaining the flexibility to choose from a variety of estimation methods. We matched on exact values where feasible (e.g., five-digit or three-digit zip code, image count) and used the coarsened matching algorithm for cases where exact matching resulted in too much discarded data (e.g., continuous image ratings for photo quality, house quality, cleanliness, attractiveness, how well-maintained, and how well-decorated). Across models 43-46, we found consistent results (Table K3_1).

However, this comes at the cost of a reduced sample size, as matching relies on overlapping matched samples with varying independent variable levels, and the data is necessarily “pruned” through this

preprocessing (Iacus, King, and Porro 2012). As can be seen from Table K3_1, matching on an increasing number of variables and more precise location indicators resulted in sharply reduced sample sizes, so the estimates may not reflect the average result for the full sample. The double machine learning analyses reported in Web App. K2, however, conducts inference on the full sample: when no overlap is identified within certain groups, the non-parametric machine learning method is capable of extrapolation, instead of tossing away observations without overlap completely.

TABLE K3_1. COARSENEDED EXACT MATCHING RESULTS

	Coarsened Exact Matching			
	(43) zip5 + Image	(44) zip5 + Image + Property	(45) zip3 + Image	(46) zip3 + Image + Property
Coefficient of Plant Image Count	0.0014*** (.0003)	0.0025*** (.0006)	0.0018*** (.0006)	0.0021*** (.0005)
Number of Matched Groups	3366	2109	3585	2282
Observations	13369	6604	15265	7427

Note: (1) For all the models, we used the log-transformed price as the dependent variable with the same set of covariates as our Model (5); (2) Values without parentheses are coefficient estimates, and values with parentheses are standard errors of coefficient estimates; (3) $0.05 < p < 0.10$, $0.01 < p < 0.05$, $0.001 < p < 0.01$, $p < 0.001$; (4) We performed the coarsening in Stata (Blackwell et al. 2009), using default Sturges bin sizes (based on the range of the data); (5) zip5 represents 5-digit zip code, zip3 or zip-3 represents the first three digits of zip code; we did exact match on zip5 and zip3.

REFERENCES FOR WEB APPENDIX

- Blackwell M, Iacus S, King G, Porro G (2009) Cem: Coarsened Exact Matching in Stata. *Stata Journal* 9(4):524–546.
- Busenbark JR, Yoon H, Gamache DL, Withers MC (2022) Omitted Variable Bias: Examining Management Research With the Impact Threshold of a Confounding Variable (ITCV). *Journal of Management* 48(1):17–48.
- Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W (2017) Double/Debiased/Neyman Machine Learning of Treatment Effects. *American Economic Review* 107(5):261–265.
- Feng Y, Liu Y, Yang S, Cai W, Zhang J, Zhan Q, Huang Z, et al. (2025) Vision-Language Model for Object Detection and Segmentation: A Review and Evaluation. (April 13) <http://arxiv.org/abs/2504.09480>.
- Frank KA (2000) Impact of A Confounding Variable on A Regression Coefficient. *Sociological Methods and Research* 29(2):147–194.
- Google DeepMind (2025) Gemini 2.5 Flash. Retrieved <https://deepmind.google/models/gemini/flash/>.
- Iacus SM, King G, Porro G (2012) Causal Inference Without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20(1):1–24.
- Knaus MC (2021) A Double Machine Learning Approach to Estimate the Effects of Musical Practice on Student’s Skills. *Journal of the Royal Statistical Society. Series A: Statistics in Society* 184(1):282–300.
- Li Y, Xie Y (2020) Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement. *Journal of Marketing Research* 57(1):1–19.
- Liu L, Dzyabura D, Mizik N (2020) Visual Listening In: Extracting Brand Image Portrayed on Social Media. *Marketing Science* 39(4):669–686.

- Minderer M, Gritsenko A, Houlsby N (2023) Scaling Open-Vocabulary Object Detection. *Advances in Neural Information Processing Systems* 36:72983–73007.
- Minderer M, Gritsenko A, Stone A, Neumann M, Weissenborn D, Dosovitskiy A, Mahendran A, et al. (2022) Simple Open-Vocabulary Object Detection. Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, eds. *Computer Vision – ECCV 2022* (Springer Nature Switzerland, Cham), 728–755.
- OpenAI (2025) Introducing GPT-5. Retrieved <https://openai.com/index/introducing-gpt-5/>.
- Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, et al. (2021) Learning Transferable Visual Models From Natural Language Supervision. *Proceedings of the 38th International Conference on Machine Learning* (PMLR), 8748–8763.
- Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(6):1137–1149.
- Scherer D, Müller A, Behnke S (2010) Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. *International Conference on Artificial Neural Networks* 92–101.
- Schuhmann C, Vencu R, Beaumont R, Kaczmarczyk R, Mullis C, Katta A, Coombes T, Jitsev J, Komatsuzaki A (2021) LAION-400M: Open Dataset of CLIP-Filtered 400 Million Image-Text Pairs. (November 3) <http://arxiv.org/abs/2111.02114>.
- Simonsohn U, Simmons JP, Nelson LD (2020) Specification Curve Analysis. *Nature Human Behaviour* 4(11):1208–1214.
- Simonyan K, Zisserman A (2015) Very Deep Convolutional Networks for Large-scale Image Recognition. Bengio Y, LeCun Y, eds. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (San Diego, CA), 1–14.
- Torrats-Espinosa G (2021) Using machine learning to estimate the effect of racial segregation on COVID-19 mortality in the United States. *Proceedings of the National Academy of Sciences of the United States of America* 118(7):1–21.
- Xu R, Frank KA, Maroulis SJ, Rosenberg JM (2019) konfound: Command to Quantify Robustness of Causal Inferences. *Stata Journal* 19(3):523–550.
- Zeiler MD, Fergus R (2014) Visualizing and Understanding Convolutional Networks. *European Conference on Computer Vision* 818–833.
- Zhang M, Luo L (2023) Can Consumer-Posted Photos Serve as a Leading Indicator of Restaurant Survival? Evidence from Yelp. *Management Science* 69(1):25–50.