UNDERSTANDING IMAGE QUALITY AND TRUST IN PEER-TO-PEER MARKETPLACES

Xiao Ma[1] Lina Mezghani[2*] Kimberly Wilber[3*] Hui Hong[4] Robinson Piramuthu[4]
Mor Naaman[1] Serge Belongie[1]

[1] Cornell Tech
[2] École Polytechnique
[3] Google Research
[4] eBay, Inc.
* Work done while at Cornell Tech

CORNELL TECH
ÉCOLE POLYTECHNIQUE UNIVERSITÉ PARIS-SACLAY
Google AI
ebay

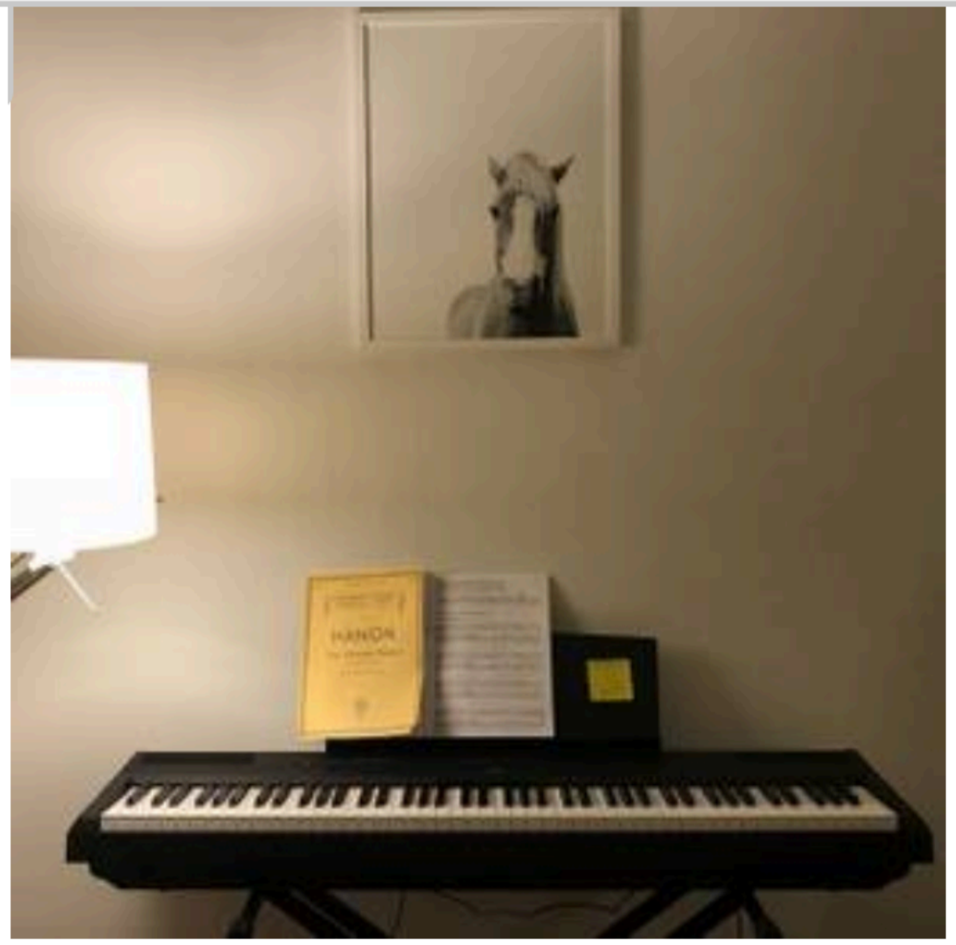
@infoxiao | maxiao.info

CONSIDER THE FOLLOWING SCENARIO

craigslist

A TALE OF THREE LISTINGS

Nov 27 CASIO WK7600 76 Key ARRANGER 4 mo. old MINT!! $350 (hvn > Orange)



Nov 25 Digital Piano -Yamaha P115 Piano 88 Key $480 (Jersey City)

Nov 28 Perfect Yamaha P255 Digital Piano w/ Bench, Stand/Pedals & Headphones $950 (Brooklyn/Bushwick)

IMAGES PLAY A CENTRAL ROLE IN MANY MARKETPLACES

Lodging (e.g., Airbnb)

Units with verified photos (taken by Airbnb's photographers) generate additional revenue of $2,521 per year on average.

For an average Airbnb property (booked for 21.057% of the days per month), this corresponds to **17.51% increase** in demand due to verified photos.

Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2017). How Much Is an Image Worth? Airbnb Property Demand Estimation Leveraging Large Scale Image Analytics.

IMAGES PLAY A CENTRAL ROLE IN MANY MARKETPLACES

Dating (e.g., Hinge)

https://medium.com/@Hinge/hinge-the-relationship-app-28f1000d5e76

CORNELL TECH · l'X ÉCOLE POLYTECHNIQUE UNIVERSITÉ PARIS-SACLAY · Google AI · ebay

RESEARCH QUESTIONS

RQ1: Can human raters **reliably judge** the quality of marketplace images?

RQ2: Can we build models to **reliably predict** high v.s. low quality marketplace images?

RQ3: What **characteristics** make high quality marketplace images?

RQ4: Does image quality affect marketplace **outcomes**?

OUTLINE

- Datasets
- Annotating Image Quality
- Modeling Image Quality
- Marketplace Outcomes
- Design Implications

SUMMARY OF RESULTS

- We created **a dataset** of real marketplace images (**≈25,000 images**) with reliable human-rated quality labels
- We were able to model and **predict image quality** with decent accuracy (**≈87%**).
- We showed that predicted image quality is **associated with higher likelihood of sales** through collaboration with eBay
- Through user experiment, we also showed that high quality user-generated marketplace images selected by our models **outperform stock imagery** in eliciting perceptions of trust from users

OUTLINE

- Datasets
- Annotating Image Quality
- Modeling Image Quality
- Marketplace Outcomes
- Design Implications

OUTLINE

Datasets

Annotating Image Quality

Modeling Image Quality

Marketplace Outcomes

Design Implications

CORNELL TECH

ÉCOLE POLYTECHNIQUE UNIVERSITÉ PARIS-SACLAY

Google AI

ebay

DATASETS

Public

letgo

Shoes: ~12,000
Handbags: ~12,000

Annotated with image quality labels

Private

ebay

Shoes: ~132,000
Handbags: ~32,000

With information associated with views and sales

OUTLINE

Datasets

Annotating Image Quality

Modeling Image Quality

Marketplace Outcomes

Design Implications

ANNOTATING IMAGE QUALITY

1. Pilot

- 50 images per batch
- 3 annotators per batch
- Rate each image from 1 (not appealing) to 5 (appealing)
- Open-ended questions to monitor task understanding

2. Label

- ~20,000 images per product category

ANNOTATING IMAGE QUALITY

1. Pilot

- 50 images per batch
- 3 annotators per batch
- Rate each image from 1 (not appealing) to 5 (appealing)
- Open-ended questions to monitor task understanding

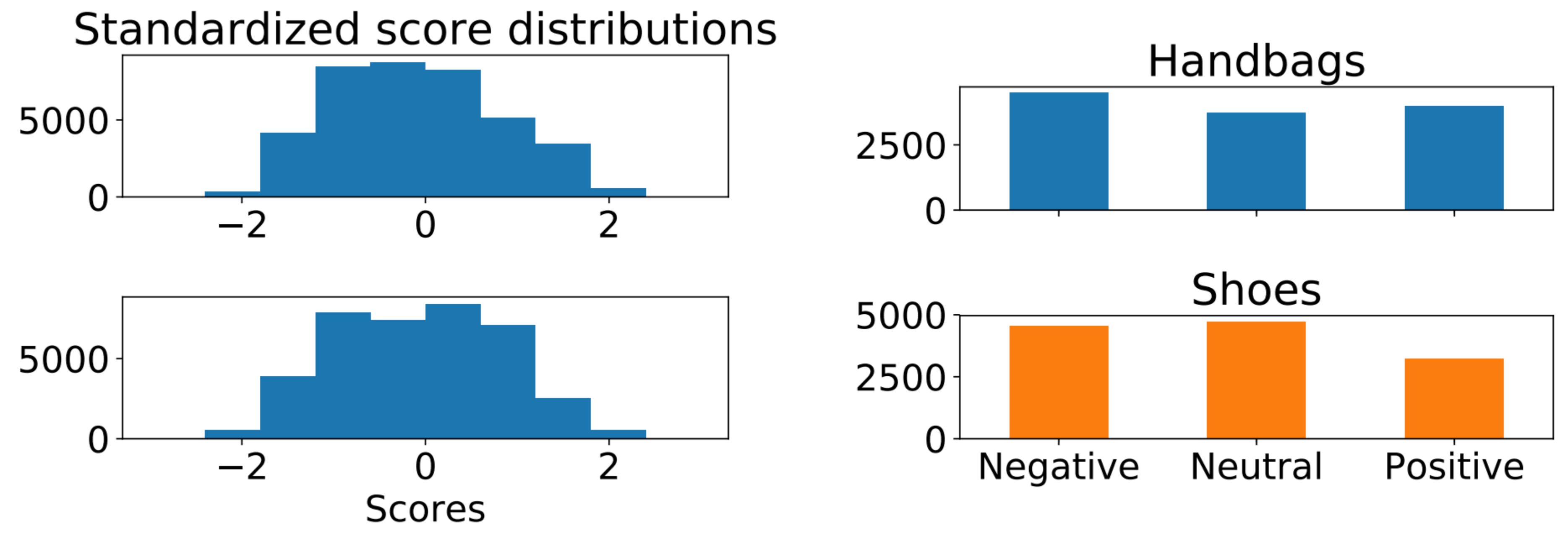
2. Label

- ~20,000 images per product category

3. Filter

- Standardize scores per rater
- Filter out images with high standard deviation across raters
- Average pairwise Pearson's: 0.70

4. Discretize



CORNELL TECH · ÉCOLE POLYTECHNIQUE UNIVERSITÉ PARIS-SACLAY · Google AI · ebay

OUTLINE

Datasets

Annotating Image Quality

Modeling Image Quality

- Prediction
- Understanding

Marketplace Outcomes

Design Implications

MODELING IMAGE QUALITY

Prediction

Model

- Fine-tuned a pre-trained Inception v3 network architecture provided by PyTorch, after removing the last fully connected layer and replacing it with a linear map down to 3 output dimensions (bad, neutral, good).
- Label smoothing: uncertainty in the data

Evaluation

- "forced-choice" — removing neutral output
- By this metric, our best shoe model achieved 84.34% accuracy and our best handbag model achieved 89.53%. (outperforms an aesthetic quality baseline model fine tuned on AVA dataset — 68.8%, and 78.8%)

MODELING IMAGE QUALITY

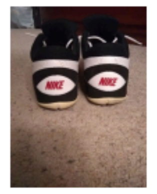
Understanding: qualitative analysis of product photography tutorials

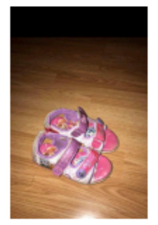
- Background (mentioned in 57% of the tutorials): white, clean, uncluttered
- Lighting (57%): soft, good, bright
- Angles (40%): multiple angles, front, back, top, bottom, details
- Context (29%): in use
- Focus (22%): sharp, high resolution
- Post-Production (22%): white balance, lighting, exposure
- Crop (14%): zoom, scale

MODELING IMAGE QUALITY

Understanding: extracting corresponding features computationally

Feature Name	Definition	Low	High
Global Features:			
brightness	0.3R + 0.6G + 0.1B		
contrast	Michelson contrast		
dynamic_range	grayscale (max - min)		
width	the width of the photo in px		
height	the height of the photo in px		
resolution	width * height / 10^6		

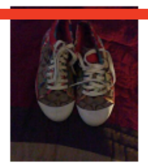
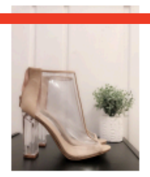
Feature Name	Definition	Low	High
Object Features:			
object_cnt	# of objects detected		
top_space	bounding box top to top of image in px		
bottom_space	bounding box bottom to bottom of image		
left_space	bounding box left to left of image		
right_space	bounding box right to right of image		
x_asymmetry	abs(right_space - left_space)/width		
y_asymmetry	abs(top_space - bottom_space)/height		

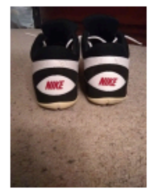
Feature Name	Definition	Low	High
Regional Features: *(fg: foreground; bg: background)*			
fgbg_area_ratio	# pixels in fg / bg		
bgfg_brightness_diff	brightness of bg - fg		
bgfg_contrast_diff	contrast of bg - fg		
bg_lightness	RGB distance from a pure white image		
bg_nonuniformity	standard deviation of bg pixels in grayscale		

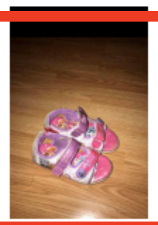
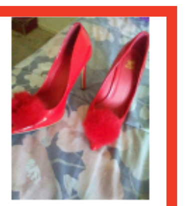
CORNELL TECH · l'X ÉCOLE POLYTECHNIQUE UNIVERSITÉ PARIS-SACLAY · Google AI · ebay

MODELING IMAGE QUALITY

Understanding: ordered logistic regression predicting image quality

Feature Name	Definition	Low	High
Global Features:			
brightness	0.3R + 0.6G + 0.1B		
contrast	Michelson contrast		
dynamic_range	grayscale (max - min)		
width	the width of the photo in px		
height	the height of the photo in px		
resolution	width * height / 10^6		

Feature Name	Definition	Low	High
Object Features:			
object_cnt	# of objects detected		
top_space	bounding box top to top of image in px		
bottom_space	bounding box bottom to bottom of image		
left_space	bounding box left to left of image		
right_space	bounding box right to right of image		
x_asymmetry	abs(right_space - left_space)/width		
y_asymmetry	abs(top_space - bottom_space)/height		

Feature Name	Definition	Low	High
Regional Features: *(fg: foreground; bg: background)*			
fgbg_area_ratio	# pixels in fg / bg		
bgfg_brightness_diff	brightness of bg - fg		
bgfg_contrast_diff	contrast of bg - fg		
bg_lightness	RGB distance from a pure white image		
bg_nonuniformity	standard deviation of bg pixels in grayscale		

OUTLINE

Datasets

Annotating Image Quality

Modeling Image Quality

Marketplace Outcomes

- Sales
- Perceived Trustworthiness

Design Implications

CORNELL TECH

l'X ÉCOLE POLYTECHNIQUE UNIVERSITÉ PARIS-SACLAY

Google AI

ebay

MARKETPLACE OUTCOMES

Sales

- We **predict** the image quality of the main eBay listing image using model trained on annotated data
- We conduct **logistic regression** controlling for number of days the listing has been on market, the number of views, and price
- Image quality predicted by our models is **associated with higher likelihood that an item is sold** (1.17x more for shoes, and 1.25x more for handbags)

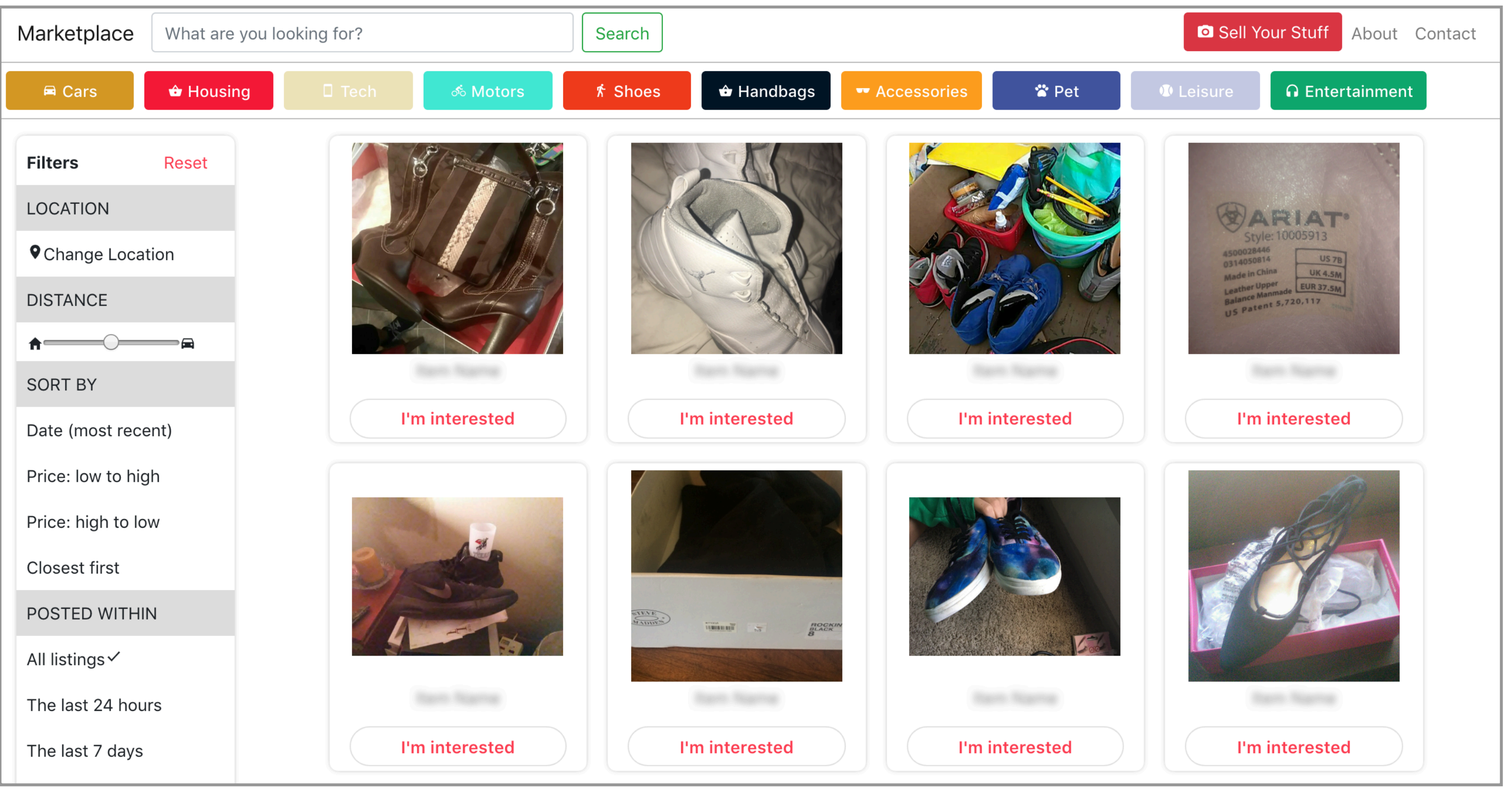
MARKETPLACE OUTCOMES

Perceived Trustworthiness: three conditions

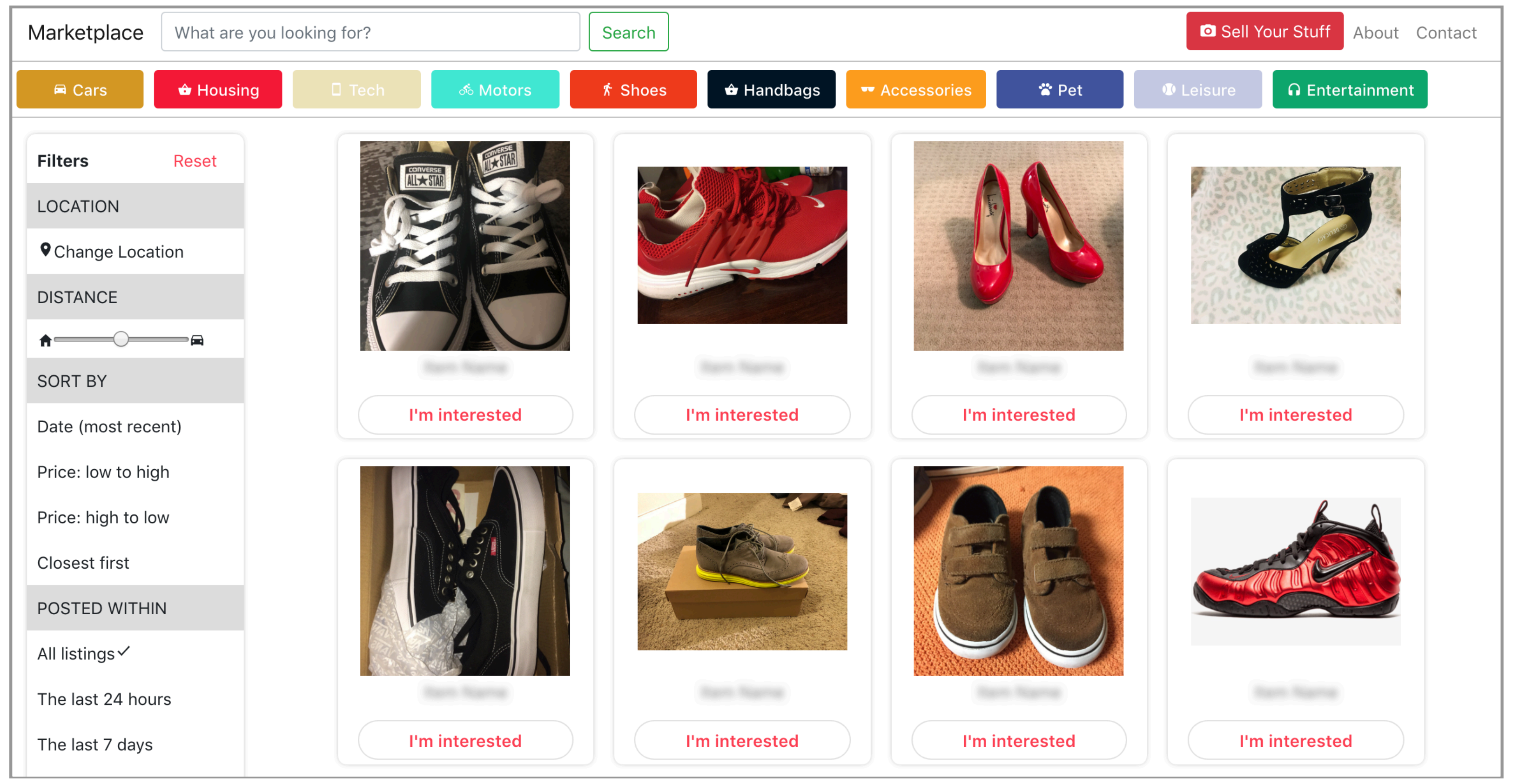
@infoxiao

MARKETPLACE OUTCOMES

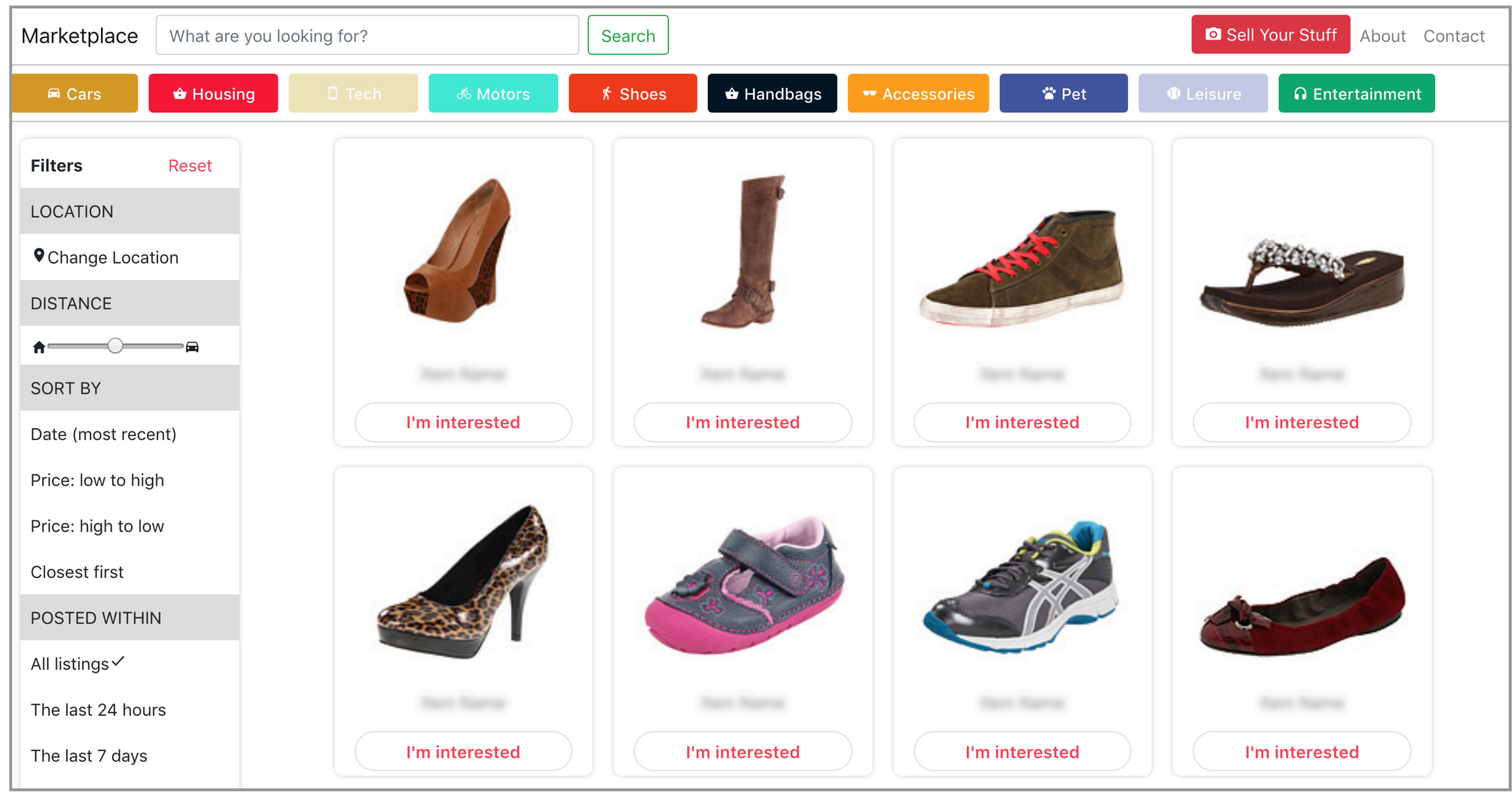
Perceived Trustworthiness: three conditions



Poor quality (predicted)



Good quality (predicted)



Stock images

CORNELL TECH · ÉCOLE POLYTECHNIQUE UNIVERSITÉ PARIS-SACLAY · Google AI · ebay

MARKETPLACE OUTCOMES

Perceived Trustworthiness: results

MARKETPLACE OUTCOMES

Perceived Trustworthiness: results

SUMMARY OF RESULTS

- We created **a dataset** of real marketplace images (**≈25,000 images**) with reliable human-rated quality labels
- We were able to model and **predict image quality** with decent accuracy (**≈87%**).
- We showed that predicted image quality is **associated with higher likelihood of sales** through collaboration with eBay
- Through user experiment, we also showed that high quality user-generated marketplace images selected by our models **outperform stock imagery** in eliciting perceptions of trust from users

LIMITATIONS AND FUTURE WORK

- Limited to two product categories
- One type of marketplace (buy-and-sell)
- Potential bias in quality prediction (especially involving faces)

OUTLINE

Datasets

Annotating Image Quality

Modeling Image Quality

Marketplace Outcomes

Design Implications

DESIGN IMPLICATIONS

Prediction-based

- Listing ranking in online marketplaces
- Automatic selection of thumbnail images

Understanding-based

- Real-time in-camera feedback to take better product photos
- Design for high-quality user-user-grated images instead of stock photos

THANK YOU

Xiao Ma[1] Lina Mezghani[2*] Kimberly Wilber[3*] Hui Hong[4] Robinson Piramuthu[4]
Mor Naaman[1] Serge Belongie[1]

UNDERSTANDING IMAGE QUALITY AND TRUST IN PEER-TO-PEER MARKETPLACES

[1] Cornell Tech
[2] École Polytechnique
[3] Google Research
[4] eBay, Inc.
* Work done while at Cornell Tech

@infoxiao | maxiao.info